AWARD NUMBER: W81XWH-11-C-0081

TITLE: Biosurveillance in a Highly Mobile Population – Year 3

PRINCIPAL INVESTIGATOR:
Nicholas A. CerJanic, MS

CONTRACTING ORGANIZATION:
QinetiQ North America, Inc.
Las Vegas, NV
89119-5272

REPORT DATE: July 2012

TYPE OF REPORT: Final

PREPARED FOR:
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*
July 2012

2. REPORT TYPE
Final

3. DATES COVERED *(From - To)*
1April2011 - 30June2012

4. TITLE AND SUBTITLE
Biosurveillance in a Highly Mobile Population – Year 3

5a. CONTRACT NUMBER
W81XWH-11-C-0081

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
CerJanic, Nicholas A.
Osterhoudt, Henry
Cochran, Christopher
Email: nick.cerjanic@apogen.com

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
QinetiQ North America, Inc.
4045 South Spencer Street, Suite 402
Las Vegas, NV 89119-5272

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)
TATRC

11. SPONSORING/MONITORING AGENCY REPORT NUMBER

12. DISTRIBUTION AVAILABILITY STATEMENT
Approved for release, distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT
The project objectives are to merge population mobility functions with an outbreak detection and characterization capability. The main areas of focus include development of techniques and technology to represent travel modes to and from the study community, and working with local stakeholders to establish the needed information sources within the healthcare, transportation, and hospitality industries. The community survey element of the research includes negotiating access to necessary and minimal datasets and documenting issues and potential impediments that must be addressed to enable such access.
The development and integration of mobility capabilities are fundamental functional requirements for a fully operational biosurveillance system, and by extension, for effective epidemiology in the computer age. This program is progressing in the effort to develop, demonstrate, and validate such functionality.

15. SUBJECT TERMS
Biosurveillance, Epidemiology, Health Informatics

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
			U	64	USAMRMC
a. REPORT U	b. ABSTRACT U	c. THIS PAGE U			19b. TELEPONE NUMBER (*Include area code)*

Table of Contents

1.0 Introduction

The project was proposed and conducted to identify and evaluate methods for integration of population dynamics with biosurveillance detection and characterization functions. The study included investigation of existing biosurveillance capabilities and available software codes as proposed to establish a point of departure and relative baseline of functional performance. The project progressed to develop the proposed predictive modeling capability and to obtain data and prepare test codes for measurement of performance improvement, if any, to be realized through the integration of regional population dynamics.

Significant efforts include negotiating with local healthcare, transportation and hospitality industry stakeholders to secure the needed information sources, and the development of detection software codes and predictive models. The project has established interface agreements and obtained and integrated data needed for situational awareness from members of the hospitality industry, from transportation industry sources, and from health care providers.

Several hypotheses were investigated as related to the project objectives. A conceptual paraphrase of the hypotheses under test is that situational awareness and response can be improved by the integration of population and population mobility information with health monitoring and tracking functions. This research focused on investigating methods and technologies potentially useful to mitigate impacts of pandemic disease or bio-weapon attack focusing on promising information integration, signal improvement, and noise reduction concepts. Based upon our review of the literature the project is unique in the direct application of population dynamics to biosurveillance codes. The study has made progress in developing and testing models and developing and testing algorithms and codes to improve representation of population dynamics in outbreak modeling and surveillance.

The project was planned to leverage the unique combination of characteristics of Las Vegas, Nevada. Factors of importance include the tourism based economy, the geographic features limiting surface travel points of egress and ingress, and the spatial concentration of visitors along a four mile strip of road. The project was undertaken in partnership with the University of Nevada, Las Vegas (UNLV) who provided essential experience and credentials in epidemiology, the required Institutional Review Board (IRB), and both credibility and trust relationships for the health care community outreach effort.

2.0 Body

Federal biosurveillance research investments since the 2001 Anthrax attacks are sizable but a fully operational capability has not been achieved. The main factors limiting progress are legislative, but technical advancements are also needed. Until data ownership decisions are made and codified and until the public is convinced individual privacy will be ensured, access to a stronger outbreak signal is unlikely.

Solutions are further confounded by the nature of health care information technology business competition. Systems and their data are non-standard because there is no mandate for standardization. There is also no incentive for standardization. Standards are not implemented due to vendor need for product discriminators and the non-universal yet continuing practice of retaining customers by ensuring the cost of changing processes and reshaping data are prohibitive. Health data shaping costs have been leveraged to create barriers to market penetration by potential competitors. These data issues along with legitimate privacy concerns and the lack of mandated standards of reporting and recordkeeping result in a very poor signal in a very noisy environment.

Due to the poor quality of the data much outbreak surveillance research, and existing applications for monitoring and reporting are focused on the sparse data problem, background noise, and selective and sensitive methods to reduce false signals yet ensure a true signal is not missed.

Meaningful integration of travel and infectious disease propagation information is highly applicable to effective epidemiology. As an awareness of the course and speed of a threat is essential to targeted intervention, an understanding of the course and speed of disease transmission is needed for complete characterization and optimal intervention during an outbreak or attack. The development and integration of surveillance with population dynamics, especially travel, should be considered essential function for effective epidemiology in the computer age.

The geography, demographics, relative centralization, transportation infrastructure, and highly refined tourism-based business focus have combined to make Las Vegas, Nevada a very suitable locale of interest for this research. Software tools have been prepared and tested which allow evaluation of the likelihood and timing of the spread of disease from an outbreak in Las Vegas to another city with emphasis on the projection of the spread of infection via surface and air travel.

The project was planned to leverage some existing technologies and add value with the development of new capabilities for: inter-city air and road travel modeling; intra-city travel and activity modeling, and; extended threat characterization to include the relationship between population movement patterns and infectious disease predictive modeling.

2.1 Background

Communicable disease models necessarily include factors provided to represent the aggregate state of understanding related to the disease, yet Hethcote (1976) writes demographic, social, cultural, and geographic factors must also be involved. Apostolopoulos and Sömez (2007) state the transportation infrastructure has made humans the most effective vector of infectious pathogens. Although there is motivation to characterize the factors influencing transmission, there is limited treatment in the literature regarding the integration of population dynamics with biosurveillance. Could the daily and seasonal human population variance noise mask an otherwise detectable signal? Can modeling mobility of and interaction between infectious and susceptible individuals provide increased utility for intervention planning? The literature indicates many surveillance systems consider spatial information to improve detection timeliness, specificity, and/or sensitivity. The integration of human population dynamics and biosurveillance is enabled by the advancement of technology and provides an opportunity-rich tested for the impact upon infectious disease modeling, biosurveillance, and public health. Kulldorf et al (2005) assessed the value of geographic information to enable focus on time series anomalies in consideration of proximity. Models with consideration of population dynamics have also been studied but validation is challenging (Busenberg and Driessche, 1990), (Sattenspiel and Dietz, 1995), (Ma and Li, 2009), (Wagner, et al, 2006).

Global models of disease spread patterns using air travel data have been prepared and evaluated including Rvachev and Longini (1985) and Grais (2002). Hufnagel et al (2004) validated a forecast capability using data from a global outbreak of severe acute respiratory syndrome (SARS) which occurred in 2003. Cooper et al (2006) stated their results argue air travel restrictions are impractical and would have little effect in delaying pandemic influenza due to the short serial interval. Sattenspiel and Dietz (1995) integrated a regional mover-stayer, migration model with a Susceptible-Infectious-Removed (SIR) compartment disease model. In addition to metapopulation level simulators some individual-level modeling has also been conducted. Elveback and colleagues (1971) prepared individual-level micro simulation models which enabled modeling of variance within the human population such as contact and transmission heterogeneity.

In air travel, factors such as proximity of passengers, length of time of travel, susceptibility of passengers and virulence of disease affect the transmission of virus from person to person. Even though the exchange of micro-organisms in pressurized cabin areas have been found to be lower than typical urban environments, the risk of exposure increases as time spent in air travel (Wenzel, 1996). Recommendations to control epidemic spreads by imposing travel restrictions, particularly for pandemic illnesses, must consider financial impact (Epstein, et al, 2007) and yet cost of intervention ceases to be a factor once a sufficiently virulent infection begins to spread.

While the concern about cross contamination among airline passengers is important, ultimately, the potential of exposed passengers and infected passengers to contaminate local populations is a public health concern. Much interest regarding the spread of disease as a result of airline travel has focused on progression of transmitting disease from one geographic area to another. Grais, et al (2004) modeled influenza forecasting based on air travel between specific American cities using data from the Centers for Disease Control and Prevention (CDC) and air traffic data from the Department of Transportation to predict outbreaks between specified large cities. Their findings indicated inconsistencies in their predictive modeling and recommended the utilization of their models as approximations of forecasting (Grais, et al, 2004). A study of the H3N2 flu virus documented the pattern of global circulation of the disease from east and Southeast Asia (Russell, et al, 2008).

At least one study found the transmission of influenza appears to be more closely correlated to air transportation flows rather than related to climate factors (Crepey and Barthelemy, 2007). Seasonal application of surveillance activities can also relate to airline travel. In the United States, influenza seasons are documented beginning October 1 of each year and are tracked for approximately 20 weeks, typically through mid-May (CDC, 2008). Research of airline transportation of the illness found that the rate of increased air transportation surrounding the Thanksgiving holiday serves as a modest predictor of influenza spread (Brownstein, Wolfe, and Mandl, 2006). However, a literature review reveals little about the effects of specific travel patterns on the spread of infection or on ways to improve surveillance through consideration of population dynamics.

The project proposed to include the use of regional demographics, transient population characteristics, tourism statistics, transportation data, and health and environmental monitoring data to develop the necessary information technologies and resulting prototype capable of modeling the spread of infection in a transient population. Timely threat containment must be the ultimate goal of surveillance therefore this demonstration project was proposed to investigate methods and develop related software to support improved intervention. Efforts included the work to define and validate functional and data requirements and to identify and assess the value of the available related datasets. The goals of the project were proposed to test and demonstrate the models and detection and characterization capabilities.

The project objectives include study of techniques and technology to represent travel modes to and from the Las Vegas study community, integration of population dynamics with existing biosurveillance methods, and working with local healthcare, transportation and hospitality industry stakeholders to establish the needed information sources. The community survey component of the research includes negotiating access to datasets and documenting issues and potential challenges to access. The project has made significant progress in obtaining, analyzing, and staging data, surveying data access issues, and in preparing software for the modeling and integration of travel functions with health surveillance.

This project leverages the unique characteristics of southern Nevada to study methods and develop capabilities useful to mitigate the effects of bio-weapons or pandemic disease. During previous efforts integration and tracking functions used semi-synthetic data, and regional and national summary data based on actual historic influenza-like-illness (ILI) summary reports to CDC, tourism, and air and road travel data. These historic temporal data for ILI, air travel, road traffic, and visitors were also used to support the investigation of algorithms for probabilistic modeling of transmission routes and patterns and to support demonstration system development and validation while awaiting actual provider data access.

The research team investigated methods, information, and processing tools with potential to provide stakeholders with an understanding of the route and pace of transmission and functions to support intervention decision-making. The integration of a travel model with detection and characterization functions is being studied to determine the advantages and complexities. The project has undertaken the tasks of development and integration of travel functions in parallel with the study of health, visit, and travel information availability and quality.

These discussions were conducted in parallel with prototype development and demonstration-database development activities, and were necessary to enable the completion of representative datasets for system validation.

The data availability and quality study supports data synthesis and assessment of signal and noise characteristics. System and study design included the information and processing for detection, travel, information integration, and intervention planning with an emphasis on projection of the spread of infection through surface and air travel. This data was staged for use in both system demonstration and validation and for use in simulation and scenario evaluation.

A visitor population individual-level travel model was prepared, integrated and outputs evaluated. Originally hosted on a dual processor single computer, the individual-level predictive modeling codes were modified to run on a Hadoop cluster of twelve workstations (from surplus on another project). This resulted in performance improvement reducing simulation processing time significantly. This cluster was later moved to a set of five T110 Dell servers resulting in additional processing time reduction.

The contact rate study was conducted first for the visitors in various behavior demographics. Later the contact rate study was expanded to resident worker and visitor interaction including surveys of local strip businesses and conventions. This empirical study was needed to gain insights into factors affecting transmission.

Codes were prepared for testing biosurveillance functions of detection and characterization with an emphasis on measurement of sensitivity, selectivity, and timeliness. Both univariate CUSUM and EWMA codes and multivariate MCUSUM and MEWMA process control codes were prepared for testing. These codes are currently being used for testing with syndromic time series data from five local hospitals over a five year timespan. Tests are being conducted and planned for all presenting, visitors only, residents only both unfiltered, parsed data and with pre-filtering. The plan includes testing of population and seasonal filters separately for comparison and in combination and evaluation of filter effects on outbreak detection.

2.2 Literature Review

Population figures based on public records and census are fixed values reflecting the number of people residing in an area. Actual daily population of a city or county varies based on resident travel, migration, visitors, commuters, birthrate, and mortality. These dynamics complicate the mathematical representation of infectious disease transmission. However, without such consideration the models of infectious disease transmission are incomplete. Korotayev (2006) offers encouragement noting that complex and chaotic behavior can be suitably represented at the macro-level by simple equations representing micro-level dynamics. This concept is applied to modeling as one seeks to represent system macro-dynamics by sufficiently modeling individual micro-level actions. Modeling when empirical data is incomplete due to business practice, privacy, competition, regulatory requirements, or resource constraints requires assumptions which in turn confound model validation (Camitz, 2010).

Much research using time-series detection methods relied on single variable approaches to obtain balance between speed and accuracy. Attempts to improve detection timeliness without excessive false positives have led to the monitoring of more than one signal, which greatly reduces both the chance of missing an alarm and the likelihood of a false alarm (Wagner et al, 2006). Evaluating a sliding time-window proved useful, but it became obvious that signal proximity had to be considered. This led to the study of algorithms for the detection of spatial and spatio-temporal clusters (Wagner, 2006, Kulldorf, 2005). Attempts have been made to model geographic spread of disease and spatial patterns of reported cases and potentially related variables, however cross correlation with local or long distance travel has not received significant attention (Carley, et al 2004).

Modeling infectious disease requires an understanding of human behavior and activities. While the severely ill can be expected to be less mobile (Longini et al, 2004) the mildly symptomatic and even those not infected, but coincidentally symptomatic, can drive the behavior of others by something as simple as a sneeze when the public is sensitized by knowledge of an outbreak, such as the during the recent novel H1N1 pandemic. At the macro level a pandemic or a smaller outbreak can be seen as an actor influencing an entity such as a city or a convention (Anolli, 2005). The spread of an infectious disease is; therefore, impacted by social interaction both at the physical location and based on individual and group perceptions. Social interaction factors transmission rate and more study appears to be warranted to support modeling of normal, baseline behavior and altered behavior.

Magnusson (2005) stressed the need for more observation based study to improve models developed using purely statistical methods. Contact rate varies substantially based on simple social activity patterns. One influential pattern is the complex movement pattern of individuals and the resulting proximity of infectious and susceptible actors. Another important pattern is the effect of information on behavior. A search of the literature reveals little study has been conducted on intra-city movement patterns, proximity, and contact rates1.

The risk of spread of disease across geographic regions has increased due to the mobility of populations. Recommendations to control epidemic spreads by imposing travel restrictions, particularly for pandemic illnesses, must take care to account for economic costs (Epstein, et al, 2007). The literature indicates most surveillance systems which consider spatial information do so only to improve detection timeliness, specificity, and/or sensitivity and do not account for population mobility. Although cross contamination is not uncommon during the transit process, spatial spread is more likely to occur once the population has reached destination points (Body et al, 2008; Ellis, Kress, and Grass, 2004; Wenzel, 1996).

Research does indicate that better tools are needed and as well as a better understanding of how the transportation network impacts the spread of disease (Hufnagel et al, 2004). They correctly note such research is essential to enable optimal intervention however, the value of travel restriction isn’t necessarily well understood. Cooper et al (2006), argue air travel restrictions may be effective for SARS, but would not work to create a useful delay in the spread of influenza. These studies reflect valuable insights concerning the potential for, and limitations of, travel-restriction interventions and indicate the costs and limited efficacy of travel restrictions, mean such drastic measures should only be taken when warranted by the severity of the threat. Other studies rely primarily on data provided by the CDC through the influenza surveillance system (Grais, et al., Brownstein). While these may be useful for developing models of transportation patterns, they do not provide the full picture of influenza and its relationship to travel.

Privacy protection issues surrounding surveillance of disease outbreaks related to hotel guests has been the subject of previous research. The European Working Group for Legionella Infections (EWGLI) created a surveillance network called the European Surveillance Scheme for Travel Associated Legionnaires Disease (EWGLINET) for reporting cases (Joseph and Rickets, 2009). This organization has been created to quickly identify and control for Legionnaires disease in the hospitality area (Cowgill et al., 2005). This European network has noted the sensitivity of the hotel industry in sharing information and has had a strict requirement for protecting privacy for clinical and travel data.

Disease outbreaks, of any size, can drastically affect a hotel and the consequences can be severe. EWGLINET was created to quickly identify and control for Legionnaires disease in the hospitality area (Cowgill et al., 2005). Once an outbreak has been detected, the accommodation site must go through a process to meet certain requirements in order to kill the disease and prevent it from spreading (Rota, Caporali & Massari, 2004). If these requirements are not met in a timely manner, the accommodation site's name will be placed on the EWGLINET's website (Rota, Caporali & Massari, 2004). In the United States, approximately 20% of reported LD cases were associated with travel (MMWR, 2007). The hope is that if clusters are detected early, the source can be quickly identified and treated. From a financial standpoint, hotels need to determine the source quickly so as to be able to return to normal business swiftly.

Transmission of influenza appears to be more closely correlated to air transportation flows rather than related to climate factors (Crepey and Barthelemy, 2007). Seasonal application of surveillance activities can also relate to airline travel. In the United States, influenza seasons are documented beginning October 1 of each year and are tracked for approximately 20 weeks, typically through mid-May (CDC, 2008). Research of airline transportation of the illness found that the rate of increased air transportation surrounding the Thanksgiving holiday serves as a modest predictor of influenza spread (Brownstein, Wolfe, and Mandl, 2006).

The 2009 H1N1 flu virus pandemic created a unique situation for modeling the spread of disease. In Mexico, especially the town of La Gloria, there began to be many cases of a respiratory illness. In La Gloria, 25% (591 cases) of the population became ill and the cause was discovered to be what became known as a novel H1N1 flu virus. Between March 10 and April 6, 591 flu cases were laboratory confirmed for H1N1 (Lopez-Cervantes et al., 2009). Cases were then found in the United States and Canada soon followed. By April 27, the first H1N1 cases in Europe were confirmed in Spain after 3 travelers returned from Mexico (Surveillance Group, 2009). In the United Kingdom, 65 cases were confirmed between April 27 and May 11 beginning with a couple returning from Mexico. France adopted an Influenza surveillance system in April after the first cases were reported around the world. By May 1, the H1N1 flu virus had arrived with travelers returning from Mexico. As of July 6, France had 358 confirmed cases with 261 of the cases attributed to travel in Mexico, the United States, Canada, South America, non-French Caribbean Islands, Asia, Oceania and the United Kingdom. The virus arrived in Greece by May 18 in a 19 year old male returning from New York City. The second and third cases were two students returning from the United Kingdom, making these cases the first to be associated with another European country. Australia and New Zealand have experienced a more severe outbreak of the virus. For the same time period, Australia and New Zealand had 8 times the amount of cases as the United States. According to the World Health

Organization (2009), there were over 6,000 deaths in 199 countries caused by the novel H1N1 outbreak by November of 2009. This is a significant increase from May 2009 when the virus had only spread to 30 countries with a confirmed 5,231 cases (Boelle, Bernillon, & Desenclos, 2009).

The ease with which this virus was able to spread poses many challenges. No country or part of the world has been immune, reinforcing the need to study the effect that travel has on the spread of disease. Flahault, Vergu and Boelle (2009) created a metapopulation model to simulate the spread of disease through 52 major cities. The state of the disease as it progresses was tracked in each city, following the four states of disease. These states are Susceptible, Exposed, Infectious and Removed (SEIR). Following their study, the authors found that there would be two major waves of the H1N1 flu virus. The first would occur in the Southern hemisphere followed by a wave in the Northern hemisphere. The tropical cities would be faced with a more moderate activity and the wave is estimated to have a longer duration (Flahault et al., 2009).

The H1N1 virus is spread as other viruses and has many of the same symptoms as the seasonal flu which includes: fever, cough, sore throat, runny or stuffy nose, headache, chills, fatigue and body aches (CDC, 2009). The CDC also reported that most of the original calculations of the virus were probably underestimated, perhaps by as high as 140 times fold (Reed, et al, 2009). Among the groups with a major under-reporting were those most susceptible to the disease, the age 5-24 population. This is significant because the upper range of that age group would include a large proportion of Army personnel including 46% of the Army's enlisted personnel and 11% of its officers fall into that age category (Department of the Army, 2005).

According to the latest information on the disease, it appears likely that an infected person can be contagious usually from one day prior to showing any symptoms to 7 days after becoming symptomatic. Importantly, contamination of animate and inanimate objects must also be taken into consideration. Based on previous studies of influenza virus, it can survive on environmental surfaces and can infect a person for 2 to 8 hours after being deposited on the surface depending somewhat upon the ambient air temperature and relative humidity.

Assumptions are often made regarding mixing, contacts, and infection when modeling infectious disease. These assumptions mean transmission is an uncertain factor (Diekmann, 1996). This uncertainty is obvious when reviewing the discourse on influenza outbreaks. What is the actual incubation period? When does an infected become infectious? Does viral shedding occur at a fixed or variable intensity? Does sunlight or humidity significantly impact susceptibility or virulence? Is there heterogeneity within the infectious population resulting in varied efficiency between those who spread the infection? Does influenza actually transmit primarily by cough or sneeze? Is a passing contact sufficient for transmission or is length of exposure also a factor? (Armbruster, 2007) (Longini, 2004) (Moser, 1979) (Kenah, 2011) (Camitz, 2010). Contact requirements are also uncertain, but evidence supports a relationship between contact rate and outbreak intensity and duration (Haber, 2007).

Much retrospective influenza epidemic analysis refers to the reproduction rate or Ro. The analysis parameter Ro is a useful assumption and simplification. Ro supports comparative evaluation of separate influenza pandemics and assessment of potentially achievable immunity levels through intervention. Ro is often called the epidemic threshold, yet also the basic reproduction number, the reproduction rate, and the reproduction number. As Ro is calculated assuming an entirely susceptible population it is a term representing the relative potential for harm. However it is only in retrospect, when the harm can be quantified Ro can be estimated.

2.3 Methodology

We proposed to investigate four hypotheses. The principal hypothesis is modeling of a highly mobile, transient population can effectively represent actual movement of people as vectors for the transmission of infectious, biologic agents. Accuracy will be measured if the resulting system can consistently detect influenza outbreaks faster than they were historically detected using conventional surveillance, and can consistently predict rate and distance of spread.

A second hypothesis is the integration of high fidelity event signals can validate the design and implementation of a time and space sensitive biosurveillance system. Once the system is validated using historic flu outbreak data, it is logical to demonstrate a realistic signal injector can effectively reproduce the results of using historic outbreak data. If the high fidelity injector consistently provides the same results for the same signal in the past, then it can be used for probability modeling, what if analysis, and decision support.

Our third proposed hypothesis is posterior probability capabilities of the validated biosurveillance system can be used to more rapidly and accurately characterize outbreaks. This is the effort to determine whether a system using temporal and spatial data as well as historic outbreak data can more rapidly detect an outbreak using posterior probability methods.

Finally, our fourth proposed hypothesis is predictive modeling using the validated biosurveillance system can support rapid threat containment. If the second hypothesis is supported by the demonstration results, then we will evaluate whether the rate and spatial distribution predictions are sufficiently accurate to support a more targeted containment strategy.

These coarsely worded statements are refined to measurable terms within the specific test.

2.3.1 Hypotheses One Evaluation

An initial test was planned to evaluate population change dynamics as a pre-filter for noise reduction in syndromic surveillance data. This test is evaluating the effect of population fluctuations on detection factors such as signal recognition, timeliness of outbreak recognition, and false outbreak signal rejection. These factors are the same as the typical measures of outbreak detection performance and are usually referred to as sensitivity, selectivity, and timeliness. As defined these terms raise sufficient questions to require interpretation. Timeliness should be simply speed of outbreak detection once an outbreak has occurred, but can be measured from occurrence of the event to detection or from data receipt to detection (Conway, 2010). Sensitivity is a term from engineering relating to the minimum signal that can be discerned and selectivity is unwanted signal rejection. However, in non-theoretical syndromic surveillance choices during primary parsing are far more influential than receiver tuning. Data cleansing, filtering, and assumptions necessary due to data inconsistency, anomalies, and ambiguities may attenuate or amplify the available and apparent signal. Choices when mapping the chief complaint to an infectious agent influence amplitude and frequency in both the signal and the background noise and syndromic data is pre-diagnosis. Selecting standard or at least often used syndrome categories has the potential to reduce this effect, but at best it is subjective analysis of subjective primary data which results in either an ideal sort of unreliable information, or, more than likely, a less than ideal one. Opportunity is presented for additional work in this area to augment study of syndrome categorization by Sholer (2004), Okhmatovshaia et al (2009), Conway et al (2010), and others.

Test of this hypothesis was proposed to include the use of existing biosurveillance algorithms and codes. The ambition, at the time of statement was measurement and contrast using accepted best practice. However naïve and ambiguous that clearly appears, the receipt of permissions to access data does provide broad opportunity for comparison with theoretical syndromic surveillance research. Consistent with that intent, this initial test and subsequent tests to evaluate population dynamics are defined to parallel and extend the research of others and contrast as possible with baseline results from prior testing. Where possible this is accomplished using the actual biosurveillance codes and information presentation developed or used by the selected previous study.

Computer-aided health surveillance based on reported syndromes depends upon algorithms to detect when rising case counts exceed a threshold indicative of an outbreak (Shmueli, 2006). Performance tests of these algorithms fill the literature, but effective comparison is challenging due to the qualities of the data. Syndromic surveillance studies using synthetic time series may include *i.i.d*. assumptions, however review of data indicates actual syndromic surveillance data typically violate such assumptions (Shmueli,2006) (Burkom, 2006).

The Multivariate Exponentially Weighted Moving Average (MEWMA) statistical process control chart tests variation in the sample mean using the exponentially weighted moving average (Lowry, 1992). An observation is compared with the mean of past observations within a time range where the moving average is calculated using weighted values. Typically values used in calculating the moving average are weighted so that the most recent observations have a greater influence on the running mean value. In manufacturing, deviations of the mean exceeding a threshold create an alarm signal to indicate an out of tolerance condition. Records of patients presenting at emergency departments (ED) can be parsed and shaped to create a time series which is somewhat similar to observed manufacturing process control data. These ED case counts vary by day, season, and situation. MEWMA charts use a sliding time window to calculate mean values and test for a condition which exceeds a selected threshold. Both above and below threshold conditions are monitored in manufacturing processes, therefore MEWMA algorithms applied to outbreak detection must be modified to be directionally constrained. Joner et al (2006) modified the MEWMA introduced by Lowry et al (1992) to be directionally sensitive.

Once the data was available and ORP approvals received the provider data was reviewed and prepared for use. Missing entries were addressed and approaches to data filtering discussed followed by test preparation. Data normalization, anomaly removal, binning of syndromes, and preliminary data analyses were conducted in preparation for test.

Also in preparation for testing, the project team evaluated some available, existing biosurveillance codes for suitability including SYDOVAT, Trisano, Real-time Outbreak Detection System, EpiFire, Global Epidemic Model and the Global Influenza Surveillance Network. However, none of these systems were selected.

The test approach for hypothesis one includes all project testing but initial tests were planned to include MCUSUM and MEWMA and univariate CUSUM and EWMA detection codes. These tests are intended to enable evaluation of the effects and benefits of separation of the visitor and resident populations for detection purposes and the effects and benefits of pre-filtering time series data with population variance and other noise-component effects. Time to signal, missed outbreaks and false positives are measured. CUSUM and EWMA codes have are coded in JAVA and MCUSUM and MEWMA codes have been prepared from MATLAB codes by porting those codes to GNU Octave.

The MEWMA codes were selected from prior biosurveillance research in keeping with the concept proposed to use existing and more importantly previously evaluated and documented capabilities. The MATLAB multivariate SPC codes were modified only as needed to port the codes to GNU Octave and to enable the selected use of either the researcher's synthesized data or empirical ED case counts from the local healthcare providers. Codes were selected from research funded by the Office of Naval Research at the Naval Postgraduate School by Fricker et al (2007) and Hu and Knitt (2007).

The MEWMA baseline is created using residuals from Burkom et al's (2006) dynamic least squares regression of the time window data which Hu and Knitt demonstrate smooths seasonal, day of the week, and holiday effects within the sliding baseline. Fricker (2007) shortened Burkom's 56 day baseline claiming optimal performance typically required windows of between 30-45 days with Burkom's 56 days as an upper limit.

This testing replaces the theoretical constructs used by Hu and Knitt (2007) with observed sample data from the five participating Las Vegas healthcare providers. his required replacement of the multivariate time series data, selected control parameters, and replacement of the prior researcher's covariance matrix with a covariance matrix calculated for the sample.

Investigation immediately reveals the contrast between theoretical synthetic data based on modulation of Gaussian white noise and actual syndromic surveillance time series data. Additionally, Fricker chose $\lambda = 0.2$ based upon observed performance and Montgomery's (2001) recommended range of $0.05 <= \lambda = > 0.25$ for the univariate EWMA. Using weight factors within the range recommended by Montgomery or at the value selected by Fricker results in false positives within the unfiltered sample. Testing with higher weights on the most recent observations reduces these false signal detections.

2.3.2 Hypotheses Two Evaluation

Evaluation of the second hypothesis employs semi-synthetic data and high-fidelity outbreak signal injection. Codes have been prepared in GNU R to produce synthetic time series and outbreaks. Preliminary tests with provider data indicate the preparation of the semi-synthetic series requires modification from prior research to preserve zip code association.

2.3.3 Hypotheses Three and Four Evaluation

Evaluation of hypotheses three and four begins with the predictive individual-level travel and infection model. Tests are in progress using historic CDC ILI data and both road and air travel data to model the paths and pace of infectious disease spread through travel. This input-output (I/O) intensive model is hosted on the cluster to leverage the Hadoop Map Reduce feature to allow parallelization of the I/O and processing.

Development of the mobility model began with the NDOT Annual Traffic Report for years 2005 through 2011. The automated traffic recorder section of the report includes a complete set of what the NDOT calls 'comprehensive summary report' pages from each of the ingress/egress routes for Las Vegas, Nevada. This information is organized by the ATR station number which is a unique identifier. Each ATR is further classified by its county, the functional classification of the roadway, and the ATR location. The Las Vegas metropolitan area can be accessed by a very limited number of major highway routes.

Typically less than half, in the past five years 43% - 47%, (GLS Research, 2008) of Las Vegas visitors travel by air. An air travel model was prepared beginning with study of the US Bureau of Transportation Statistics (BTS) (Citation needed) data available online via queries and reports. The BTS data were used to create tables of aircraft types, seating configurations, and passenger capacity for each aircraft model and configuration used by airlines serving Las Vegas McCarran International Airport, airport code LAS. This study is intentionally focused on the airports and airlines having direct flights to and from LAS. Over the twelve year timeframe coinciding with the road travel model 297 US and international airports had direct flights to or from LAS with an annual average number of 220 airports serving passengers with direct flights to or from LAS during any single year within the model.

The research team sought to identify hotels that would be willing sources of information to improve public health surveillance. We identified 19 hotel ownership companies representing 40 different properties. Of these ownership chains, the largest in order of properties owned were MGM-Mirage (12 strip properties owned on the Las Vegas strip); Harrah's Entertainment (7 properties owned on or near the Las Vegas strip); Boyd Gaming (3 properties on/near Las Vegas strip downtown; 4 Coast properties owned, 2 near LV strip and 2 off strip properties); Wynn Resorts (2 properties on Las Vegas Strip); and Sands Corporation 2. The project team interviewed security and risk management personnel and examined related artifacts to determine the types of information they collect on guests who become ill or injured, date and time of guest complaint/variance, whether they maintain this data in any storage capacity, how they respond to guests who become ill, the disposition of those guests, and both their interest and willingness to participate in the research project.

Project efforts included the development of software providing functions for air and surface mobility modeling and simulation of travel and infection in a locale of interest. Advancements in computer performance have enabled modeling of travel and disease transmission at the level of the individual traveler. Individual-level models (ILM) enable modeling of heterogeneity and variance not possible in metapopulation infection spread models. Datasets were prepared from airline flight schedules, aircraft model and seating configurations, and from Nevada Department of Transportation (NDOT) automated traffic recorders for a five year span.

Due to the large number of datasets and the size of some of those datasets the time required to process data for simulation and testing was considerable. Some work was done to improve performance by standardizing the interfaces between components. This allowed distribution of application components over multiple processors. This did improve performance but the application's performance was mainly impacted by input and output requirements during simulation operation which were not significantly mitigated by process distribution. The input-output processing issue was addressed by parallel processing and by using the Map-Reduce feature of a Hadoop cluster. Procedures for operation of the cluster are provided in Appendix B.

The ILM includes a simulator for disease or infectious agent within the regional population and allows modeling of contact and transmission heterogeneity. The disease simulator is integrated with an individual travel model by simulating persons of epidemiologic interest and their time, path, and mode of transportation. Disease or infectious agent scenario files are used to set the parameters for average disease latency, virulence, and duration of infectivity. Influenza-like-Illness (ILI) was selected as the infection for this study based on availability of syndromic data, CDC sentinel seasonal and pandemic flu outbreak histories, and the available discourse related to influenza and biosurveillance.

Following data preparation and staging, and hand optimization of codes, the processing and I/O requirements are not out of reach of a workstation cluster with cycle-execution approximately twenty minutes. Tests were conducted using a dual Xeon processor server initially requiring approximately twelve hours per cycle. A SUN V880 with four-processors and two RAID arrays was available for use and performed no better than the dual Xeon server. Excessed and fully-depreciated workstations from another Federally funded program were then assembled into a twelve-computer cluster and loaded with open-source operating systems and component open source parallel processing software. This Dell Precision 360 cluster was capable of cycle times of less than thirty minutes. However, that equipment needed to be returned under contract related regulation so five Dell T110a servers were acquired in an attempt to match the performance of the twelve workstation cluster. This five-server cluster resulted in cycle times of approximately twenty minutes.

2.3.4 Outreach and Data Collection

2.3.4.1 Provider Data

Access to data for testing required interview of stakeholders and data owners to investigate issues and constraints. Project researchers conducted a series of structured meetings with local stakeholders and visited local hospitals, clinics, and private practice physicians to investigate technical, operational, and policy issues related to surveillance information access. These outreach activities include discussion the Emergency Department data qualities and potentially useful interface protocols. Through these interactions data was obtained from five local hospitals:

- Valley Health Systems (3)
  - 2006-2007 = 110,165 visit records
  - 2008-2009 = 112,638 visit records
  - 2009-2010 = 123,450 visit records
  - 2010-2011 = 148,948 visit records

- Sunrise Hospital (1)
  - 2007 = 79,398 visit records
  - 2008 = 88,623 visit records
  - 2009 = 97,312 visit records
  - 2010 = 100,381 visit records
  - 2011 = 110,005 visit records

- University Medical Center (1)
  - 2004 = 65,534 visit records (years overlap)
  - 2005 = 53,047 visit records (years overlap)
  - 2006 = 12,867 visit records
  - 2007 = 10,080 visit records
  - 2008 = 10,197 visit records

- UMC HL7 Feed
  - ~9.9 million messages
  - ~2.7M ER, ~7.2M ADT
  - > 1 message per visit

2.3.4.2 Contact Rates

As a facet of the study to determine how disease spreads through the population of Las Vegas and especially the visiting population, it is necessary to approximate the social interactivity of individuals that frequent the Las Vegas Resort corridor. In prior work, in order to define actual contact rates to populate our Susceptible, Exposed, Infected, Recovering (SEIR) models, researchers determined rates for the most common gaming behaviors for Las Vegas visitors. This year, to further define contact rates, researchers investigated Las Vegas residents working on the Las Vegas Strip and convention attendees.

Residents Working on the Las Vegas Strip

During research to support our biosurveillance project we needed the figure of Las Vegas residents who worked on the Las Vegas Strip, The area on Las Vegas Boulevard from the stratosphere Tower on the North to Mandalay Bay on the South. Data was readily available for employees working in casinos from research done by the Center for Gaming Research at The University of Nevada Las Vegas (UNLV). That total was 120,000. The number of Las Vegans working for non-casino entities; however, was not available.

To find this number Dr. Henry Osterhoudt conducted a survey of all the businesses on the strip. The survey included: retail outlets (stores, kiosks, and mini-marts), restaurants (fast food and sit down), night clubs, valet parking, tour companies, ticket vendors, rental agencies, massage parlors, street performers, street vendors, motels , tattoo parlors, and time shares. The researcher visited 642 separate businesses. The number constitutes all the businesses on the strip including those physically located in resorts but not owned by the casino corporation. These entities rent space from the resort but are owned by a separate entity. The number includes all the businesses in the various malls along the strip: Stratosphere Tower Shops, Fashion Show Mall, The Grand Canal Shoppes at the Venetian, The Shoppes at the Palazzo, The Forum Shops, Via Bellagio Shops at Bellagio, Miracle Mile Shops at Planet Hollywood, Crystals at MGM Mirage City Center, and Mandalay Place at Mandalay Bay. In addition other casinos have groupings of shops in or adjacent to their properties, for example between Wynn and Encore or between Luxor and Excalibur. At each business the researcher asked a responsible manager or the person manning the business or kiosk how many people worked at the establishment in a 24 hour period. Some of the establishments had business hours ranging from 8 to 16 hours. Some were open 24 hours a day.

The survey took three weeks and determined that a maximum of 20,156 individuals work on the strip in non-casino owed businesses on any given 24 hour period.

Contact Rates for Convention Attendees

Researchers surveyed contact rates for convention attendees in Las Vegas. The research was done during the Consumer Electronics Show (CES) 10-13 January 2011 and during observations of smaller conventions at various resorts during the year. The CES is a huge convention staged at the 3 million square foot Las Vegas Convention Center (LVCC) which includes 2 million square feet of exhibition space and 243,000 square feet of meeting rooms and the 2.2 million square foot Venetian Convention Center. The show was attended by over 150,000 people. During the convention researchers acted as convention goers and recorded their contacts in two ways. The first set of numbers was determined by counting the total number of contacts that came within three feet of the front of the researcher. The numbers were recorded over a three day period as the researcher acted as a convention attendee arriving at the convention, registering, and then touring all the exhibits. The second set of numbers was determined as those contacts that lasted longer than three minutes. This set was determined by simulating a convention goer who was conversing with convention vendors or listening to vendor presentations. As in the research of gamers the largest numbers of contacts were accumulated during transit of the convention. Researchers recorded their contacts in 15 minute intervals from the time they exited their vehicles until they returned to their vehicles at the end of the day. Researchers were Las Vegas residents and thus not staying at a resort hotel. Contacts tallied 357 per hour although the numbers varied greatly depending on whether the researcher was actually moving about the convention or simply getting there or returning to their transportation.

The contact rate dropped markedly when the time of 3 minutes was included as a parameter. Researchers began their research by attempting to count both types of contact but quickly realized that this was extremely difficult so a separates effort was made to specifically determine the contact rate only for the three minute parameter. This contact rate was significantly smaller than the prior rate with an average of 3 to 6 per hour. Estimating the number of convention goers who experienced this contact rate was possible only by an educated observation, not an actual count. The estimate is about 15% of convention goers seemed to be in this category. But the figure could skew higher.

As with gamers the majority of contacts were experienced while traversing the convention. Choke point and popular exhibits also contributed to the larger numbers as did the huge number of attendees who taxed even the huge capacity of the LVCC. This convention was one of the largest in total attendance, but it is not out of the norm for contacts of attendees. Smaller conventions use smaller venues, but the contacts of attendees are similar. Movement and choke points at the various venues in Las Vegas, each casino resort has some convention or meeting space which accommodate various size meetings or events, are for the most part consistent in elevating contact rates. It should be noted; however, that architecture does affect contact rate to an extent. Newer convention and meeting facilities are designed with larger hallways, more spacious meeting rooms and multiple routes of ingress and egress. The sum total of these architectural advances is to decrease the contact rates for transiting conventioneers and meeting attendees. Older facilities, many of which are still in use, do not have the wider routes and more spacious venues of the newer properties. For the largest conventions which all use the LVCC convention facilities this increases the contact rate because the Las Vegas Hotel and Casino, Previously the Las Vegas Hilton is an older facility and is contiguous to the LVCC. The LVCC itself is a huge facility but it encompasses routes which constrict movement of huge convention audiences and it does not have sufficient dining venues to handle the huge crowds for the largest conventions without congestion. In fact although the LVCVA tries to alleviate the congestion as much as possible additional dining venues would not prove viable. Likewise the Sands Expo Convention Center is an older facility and it like the LVCC has its share of chokepoints even though the resorts to which it is connected, The Venetian and The Palazzo, are brand new and state of the art.

In addition the growing number of attendees at some of the more popular events; the CES is a good example, contribute to the crowding. The LVCVA attempts to alleviate this problem by expanding the convention to multiple venues at different locations. The problem is that at any convention certain exhibitors have more popular exhibits than others and these exhibits whether because of the exhibitor or the product cause conventioneers to congregate at those locations. At the CES new electronics (The LG exhibit for example) and new vehicles drew capacity shoulder to shoulder crowds. In some cases exhibitors who have exhibit space near entrances to the convention floors, space which is highly desired, also contribute to congestion as attendees crowd together to observe the displays or the interactive experience. Again savvy exhibit designers seek to grab and hold the attention of attendees and occupy the space near the entrance contribute to the congestion largely by design. These factors, despite the best efforts of the event organizers, greatly effect congestion and drives up contact rates.

Additionally at the LVCC security is tasked with admitting only authorized attendees. At each entrance security personnel check identification badges. This creates bottlenecks and further contributes to elevating contact rates as attendees queue up to enter the convention hall or go from one building to another. Each entrance has another security checkpoint and the identification process is repeated.

Conventions habitually last for a period of days which also elevates contact rates. Meeting and events which last for one day do not afford the attendees sufficient exposure time to effect an increase in contact rates so a multi-day convention is the most representative and the best laboratory in which to determine an accurate effective rate.

Most studies of disease have assumed a homogeneous contact rate instead of doing the research to accurately determine the actual rate of contacts. This study has done extensive research to provide actual data that models subject behavior. Our researchers have spent a good deal of time modeling both gamer and convention attendee behavior on the Las Vegas strip. We have used data gathered by both the Las Vegas Convention and Visitors Authority and the University of Nevada Las Vegas Center for Gaming Research to focus and refine our research. This data served as a departure point to permitting our personnel to maximize the effectiveness of our activities. For example we knew percentages of gamers who played various games so we were able to focus on behavior of gamers who played the most popular games thus providing the largest sample of visitor behavior. We also knew the size and frequency of conventions and the use of convention and meeting space so we were able to most effectively employ our researchers to acquire real contact data.

2.4 Analysis

2.4.1 Provider Data

The following report summarizes and analyses the syndromic time series data from participating providers and was prepared by Dr. Chris Cochran of UNLV.

Christopher R. Cochran, Ph.D.; Subcontractor PI
University of Nevada Las Vegas
School of Community Health Sciences

Bio-Surveillance of a Highly Mobile Population
Understanding Influenza and Influenza-like (ILI) Symptoms

Influenza is considered a seasonal illness typically spanning October 1 – Mid-May of each year. Therefore, for historical data collection purposes, annual influenza and influenza-like illnesses must be categorized in the appropriate time frame. The Centers for Disease Control and Prevention (CDC), monitors influenza from state and local health departments, federal agencies such as the Department of Defense and Veterans Affairs, and sentinel sites including physician offices, health care clinics, hospital emergency departments and urgent care facilities, and the Department of Defense and Veteran's Affairs (CDC, 2008). According to the CDC, ILI includes fever, headache, fatigue, cough, sore throat, runny or stuffy nose, body aches and diarrhea and vomiting (more common in children than adults). They note that it is impossible to diagnose flu based presence of symptoms alone because other diseases can have similar symptoms. The only way to confirm influenza is through the use of clinical testing (CDC, 2008).

It is our intent to develop a system whereby patient visits can be submitted for the project that relate to influenza like illness (ILI) on an ongoing real time or near real time basis. To develop and adequate model for understanding visitor utilization of local hospitals and providers, the project also sought to collect historic patient visit information for the previous five years. By obtaining patient zip codes as part of the data collection process, an analysis of the number of visitors utilizing health care providers can assist in developing the transportation model. This analysis will also allow us to compare how well chief complaints match up to diagnoses.

Based on four- year data trends as reported by the Nevada State Health Division, reports of ILI illness have increased significantly at the beginning of each year, typically around the 10th week of the influenza season. In Figure 1, the actual peaking of ILI begins in early December, then drops slightly during the holidays and begins to show rapid acceleration at about week 3 of the at the beginning of the year. This is notable because the Las Vegas visitor volume drops during the month of December then picks up significantly in January (LVCVA, 2008).

Data Needs

ILI typically refers to fever and one of the following: headache, cough, sore throat, runny/stuffy nose, body aches, diarrhea and vomiting. However, some symptoms may not be present during patient visit and diagnosis may reflect a more general description such as lower respiratory infection, pneumonia, or upper respiratory infection. To that end, the project needs to identify all complaints that can fall into the ILI category. For the purpose of this study the following data needs were identified:

- Pseudonymized linker (patient de-identifier measure)
- Event time and Place (for the patient encounter)
- Age. Age may be an important components since children, for example, may have different influenza like symptoms (e.g., vomiting) than adults.
- Zip code. 5-digit zip code or 3-digit for sparsely populated zips.
- Patient classification. Hospital patient classifications generally include emergency room, inpatient, outpatient, or other services such as laboratory or radiology. In this case, only emergency room classifiers are necessary since we are primarily interested in ambulatory patients. Outpatient information would typically apply only for follow-up visits. Inpatient classification may be useful, but not necessary for this project.
- Chief complaint. This is the patient reported reason for seeking care. Key for this project. Need to understand how this information is collected and coded. (See section on ICD-9 coding criteria).
- Illness onset by date/time (desirable for this study but is not routinely collected for electronic data entry). Probably would require review of physician, nursing or triage notes.
- Diagnosis/Injury code. Diagnosis or diagnoses assigned from patient visit. This is the billing code that will be the most reliable for case identification and confirmation. However, the availability of this data will vary from hospital to hospital.
- Diagnosis type (preliminary, interim, final, admitting).
- Diagnosis date/time. Should be easily available for date. May not be consistent for time.
- Discharge disposition. Essential element but may only be known as admitted to hospital, sent home, AMA, other).

To determine the locale of visitors and potential onset of their illness, other useful information would include visitor place of stay, days since arrival, and days until departure.

Data Collection and Methodology Techniques

Hospital emergency room data for the years 2006-2010 were used for this study. The data was compiled from hospitals that have the closest proximity to the Las Vegas, NV strip corridor. All hospitals included in this study are located within (X) miles of that corridor. Through interviews with local resort security operators, Southern Nevada Health District, and emergency services personnel, these hospitals were identified as having the greatest likelihood of providing emergency services to visitors residing on the strip corridor: University Medical Center, Sunrise Medical Center and Sunrise Children's Hospital, Desert Spring Medical Center, Valley Hospital and Medical Center, Spring Valley Medical Center.

An IRB from the previous study was updated and resubmitted to the UNLV Office for the Protection of Human Subjects prior to the collection and received final approval by the UNLV IRB in October of 2011. Final approval of the IRB project from the Human Subjects Protection Scientist (General Dynamics) Human Research Protection Office (HRPO), Office of Research Protections (ORP), U.S. Army Medical Research and Materiel Command (USAMRMC) was given approval in February of this year. Therefore, data collection for the project was delayed until the final approval from the sponsor agency.

Data files were transmitted through secure email files with expiration dates upon acceptance of the files from UMC and Valley Hospital. Data from Sunrise Hospital was transmitted into a CD. Data was formatted into Excel comma delimited files.

UMC has been a partner in this project since project year 1. Both UMC and Sunrise Medical Center represent the largest hospitals in Southern Nevada thus experience higher volumes of emergency room visits. UMC also operates a level one trauma center, but data from that emergency unit is not included in this analysis since it is not likely to have The data from all other hospitals was collected during the third funding year of this project. Desert Springs Medical Center, Valley Hospital and Medical Center and Spring Valley Medical Center are all part of the Valley Hospital Systems (VHS). The data collected from these hospitals was provided by their central data source. All data providers were given the data elements for the collected data. Some fields were inconsistent and one of the most important data components, "Chief Complaint", was available for only one year of the VHS data. Data was collected in an excel data delimited format.

In the period 2006 – 2010 the number of visitors to Las Vegas ranged from just over 36 million more than 39 million per year. The period 2007 to 2009 saw decreasing number of visitors to Las Vegas due primarily to the economic recession. However, in 2010 the numbers began to climb again to more than 39 million visitors, still below the averages of 41 million tourists reported in our previous study.

For this study, data was collected for a five year period from the hospitals for the period 2006-2010. The data elements considered in this study included the following:

De-identified patient code, admission date, admission time, discharge date, Chief Complaint, up to five diagnosis (ICD-9) billing codes, age, sex and patient zip code.

There are some gaps in the data that will be addressed in a follow-up report. These gaps include missing data for 2008 from the VHS hospitals and missing data from 2006 from Sunrise Hospitals. The table below illustrates the data collected from the hospitals. The data indicates that more than 15% of the ER visits to area hospitals are by visitors (see Table 1).

Table 1 – Hospital emergency room utilization by local residents and visitors

			HOSPITAL					
			UMC	SUNRISE	SPRING VALLEY	VALLEY	DESERT SPRG	Total
local	0	Count	27901	74924	32287	23310	22399	180821
		% within HOSP	8.4%	15.8%	20.5%	14.4%	21.4%	14.7%
		% of Total	2.3%	6.1%	2.6%	1.9%	1.8%	14.7%
	1	Count	302691	400438	125010	138175	82092	1048406
		% within HOSP	91.6%	84.2%	79.5%	85.6%	78.6%	85.3%
		% of Total	24.6%	32.6%	10.2%	11.2%	6.7%	85.3%
Total		Count	330592	475362	157297	161485	104491	1229227
		% within HOSP	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
		% of Total	26.9%	38.7%	12.8%	13.1%	8.5%	100.0%

The addition of the other hospital data suggests that an even greater volume of patients visit the private hospitals than visit the county's only public hospital. This may be due likely to the overcrowding of the public hospital and the insured nature of the area's visitors. But the additional data is of major importance in trying to determine the utilization of Southern Nevada hospital emergency rooms by visitors to the community. An analysis was conducted to determine the top DRG elements for the report. Based on the information provided, the following indicate the main codes billed by the hospitals (Table 2).

Table 2: : ICD Code Frequency of Visitor Utilization of Hospital ERs

University Medical Center				
Rank	ICD-9 Code	Diagnosis	Frequency	Pct.
1	789.00	Other symptoms involving abdomen and pelvis	31182	7.5
2	780.6	Fever and other physiologic disturbances of temperature regulation	15830	3.8
3	729.5	Pain in Limb	14985	3.6
4	786.2	Cough	13809	3.3
5	V71.4	Observation following other accident	13163	3.2
6	787.03	Vomiting alone	10250	2.5
7	784.0	Headache	10087	2.4
8	780.60	Fever and other physiologic disturbances of temperature regulation	9151	2.2
9	724.5	Fever and other physiologic disturbances of temperature regulation	8794	2.1
10	786.50	Chest pain	8372	[illegible]

Sunrise Hospital and Medical Center				
Rank	ICD-9 Code		Freq.	Pct.
1	V71.9	Unspecified Diagnosis	12785	2.7
2	465.9	Acute upper respiratory infections of multiple or unspecified sites	10281	2.2
3	305	Nondependent abuse of drugs	9090	1.9
4	648.93	Issues of Pregnancy	9053	1.9
5	780.6	Fever and other physiologic disturbances of temperature regulation	8518	1.8
6	786.59	Other discomfort in Chest	8408	1.8
7	786.5	Chest pain	7005	1.5
8	599	Other disorders of urethra and urinary tract	6440	1.4
9	382.9	Other symptoms involving skin and integumentary tissues	6108	1.3
10	780.2	Syncope and collapse	5965	1.3

VHS Hospitals				
Rank	ICD-9 Code		Freq.	Pct.
1	789	Other symptoms involving abdomen and pelvis	16581.0	3.1
2	305	Nondependent abuse of drugs	12758.0	2.4
3	786.59	Other discomfort in Chest	10644.0	2.0
4	786.5	Chest pain	8740.0	1.6
5	465.9	Acute Upper respiratory infection	7806	
6	780.2	Syncope and collapse	7195.0	1.3
7	599	Other disorders of urethra and urinary tract	6748.0	1.3
8	784	Symptoms involving head and neck	5758	1.1
9	V68.9	Unspecified administrative purpose	5065	0.9

*10th ranked in VHS unable to determine.

The data in the tables above illustrate one of the major problems in using ICD9 data codes for early identification of outbreaks such as flu. While the data from the UMC hospital indicates a greater likelihood of potential influenza like illness (ILI), the data from all of the other hospitals appears to be more consistent in their reporting measures. To calculate the data included in these tables, an analysis was conducted of all ICD-9 codes provided (up to 6 codes in some cases). One of the limitations of this data pertains to the Valley Health Systems hospitals which only reported on ICD-9 code for their cases. Thus, it is possible that inclusion of more than one code would have captured a truer assessment of the patient services. In examining the data from the other hospitals, the great majority of cases had more than one ICD-9 code reported, thus, it appears unlikely that the cases provided in the VHS hospitals' data would have included less than one code. It is also possible that coding errors, changes in data collection system formats, or other factors including time needed for proper data submission contributed to the lack of multiple codes in these cases.

In Table 3, we sorted the top ten ICD primary complaint code (the first billing code assigned to patients). In this table we use only the first ICD-9 code due to missing values from the VHS hospitals.

Table 3: Top ICD-9 Codes, Visitors vs. Local Residents for primary ICD-9 code

Visitors (2006-2010)				Local Residents 2006-2010			
Dx		Freq.	PCT.	DX	Code	Frequency	Percent
Nondependent abuse of drugs	305	9008	5	Unknown DX	V71.9	31180	3
Syncope and collapse	780.2	5406	3	Other symptoms involving abdomen/stomach	789	24062	2.3
Unknown DX	V71.9	3749	2.1	Other discomfort in chest	786.59	20042	1.9
Other discomfort in chest	786.59	3597	2	Other symptoms involving abdomen/stomach	789	16122	1.5
Chest pain	786.5	2848	1.6	Chest Pain	786.5	12907	1.2
Other symptoms involving abdomen/stomach	789	2657	1.5	Other disorders of urethra and urinary tract	599	12627	1.2
Symptoms in digestive sys	787.03	2445	1.4	Flu Symptoms	465.9	11789	1.1
Other gastrointitis	558.9	2263	1.3	Issues of soft tissue	729.95	11630	1.1
Other disorders of urethra and urinary tract	599	2077	1.1	Nondependent abuse of drugs	305	11542	1.1
Contusion	920	1578	0.9	Chest Pain	786.62	10613	1
Pneumonia (#12)	486	1483	0.8	Fever	780.6	10150	1
Acute sore throat NOS (#18)	462	1162	0.6	Acute sore throat (NOS) (#22)	462	6923	0.7
Flu symptoms (#24)	465.9	994	0.5		784	9998	1
Fever (#25)	780.6	940	0.5		780.2	9056	0.9

Based on the numbers in the table, the types of illness diagnosed indicate very little difference in frequency after the top 10 codes. For the visitors data, we included the code for the flu related symptoms which rank 24th on the list as well as some prominent ILI type symptoms. A complete list of these codes for up to 5 diagnostic codes will be provided in our final report.

Identifying Cases from Chief Complaints

This preliminary analysis is critical to the early detection of any cases beyond the norm. Often, a patient may present to the emergency room with full knowledge of their condition, but cases related to flu may not be so clear. When considering ILI conditions, a number of symptoms may contribute to an ultimate detection of a case. However, some cases may be vaguer. Cough, for example, is a vague symptom taken by itself because the condition may be caused by other, sometimes similar respiratory illnesses such a bronchitis or allergies. However, based on most of the literature, the combination of cough and other symptoms, especially fever, can be a good indication of flu. To ascertain the chief complaints that could more reliably be considered a chief complaint of flu, we first had to isolate specific terms in the chief complaint. Based on previous literature reviews, we selected those terms that were most likely to be used in describing symptoms of flu. The most obvious were those cases in which the chief complaint was flu or influenza. Next, we compiled cases using specific symptoms in some string of the data. Those symptoms included the following:

- COUGH
- COLD
- FEVER
- RUNNY NOSE
- WEAKNESS
- BODY ACHES
- SORE THROAT
- HEADACHE

Those codes cases were then recalculated into a binomial using 1 for the presence of the symptom and 0 if the symptom was not present. Based on those findings, we then merged data by using the following combinations (examples are shown based on the merged data sets from UMC and Valley Hospital where 1 = the presence of two or more symptoms and 0 = no ILI symptoms:

FLU

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	.00	1225522	99.7	99.7	99.7
	1.00	4055	.3	.3	100.0
	2.00	2	.0	.0	100.0
	Total	1229579	100.0	100.0	

FEVER

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	.00	1184095	96.3	96.3	96.3
	1.00	45484	3.7	3.7	100.0
	Total	1229579	100.0	100.0	

RUN_NOSE

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	.00	1227436	99.8	99.8	99.8
	1.00	2143	.2	.2	100.0
	Total	1229579	100.0	100.0	

BODY_ACHE

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	.00	1228192	99.9	99.9	99.9
	1.00	1387	.1	.1	100.0
	Total	1229579	100.0	100.0	

SORE_THT

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	.00	1219788	99.2	99.2	99.2
	1.00	9791	.8	.8	100.0
	Total	1229579	100.0	100.0	

COUGH

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	.00	1202635	97.8	97.8	97.8
	1.00	26944	2.2	2.2	100.0
	Total	1229579	100.0	100.0	

STUFFY_NS

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	.00	1229408	100.0	100.0	100.0
	1.00	171	.0	.0	100.0
	Total	1229579	100.0	100.0	

VOMITTING

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	.00	1228176	99.9	99.9	99.9
	1.00	1403	.1	.1	100.0
	Total	1229579	100.0	100.0	

Any data in the table above indicates that of 1,229, 579 cases examined, more than 91,000 hospital visits included at least one of the symptoms for ILI. Any cases resulting in a score of 2 or more could be considered the combination necessary for determining flu. The result was 2,319 cases for the two hospital systems. That data was then merged with those cases that were classified as flu or influenza:

FEV_STUFFY

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	.00	1229569	100.0	100.0	100.0
	1.00	10	.0	.0	100.0
	Total	1229579	100.0	100.0	

FEV_RUNNY

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	.00	1229399	100.0	100.0	100.0
	1.00	180	.0	.0	100.0
	Total	1229579	100.0	100.0	

FEV_COUGH

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	.00	1226649	99.8	99.8	99.8
	1.00	2930	.2	.2	100.0
	Total	1229579	100.0	100.0	

COUGH_THRT

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	.00	1229102	100.0	100.0	100.0
	1.00	477	.0	.0	100.0
	Total	1229579	100.0	100.0	

FEV_THROAT

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	.00	1228953	99.9	99.9	99.9
	1.00	626	.1	.1	100.0
	Total	1229579	100.0	100.0	

COUGH_STUFFY

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	.00	1229559	100.0	100.0	100.0
	1.00	20	.0	.0	100.0
	Total	1229579	100.0	100.0	

COUGH_RUNNY

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	.00	1229197	100.0	100.0	100.0
	1.00	382	.0	.0	100.0
	Total	1229579	100.0	100.0	

COUGH_ACHES

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	.00	1229504	100.0	100.0	100.0
	1.00	75	.0	.0	100.0
	Total	1229579	100.0	100.0	

FEV_ACHES

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	.00	1229438	100.0	100.0	100.0
	1.00	141	.0	.0	100.0
	Total	1229579	100.0	100.0	

THROAT_ACHES

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	.00	1229518	100.0	100.0	100.0
	1.00	61	.0	.0	100.0
	Total	1229579	100.0	100.0	

When combined with the flu and influenza variables, the total number of cases is approximately 2,300 cases. In the table below, the variable ILI_COMBO represents the number of ILI related cases through the merging of those variables with at least two symptoms of flu. The data indicates that 4,649 cases can be realistically classified as ILI.

ILI_COMBO

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	.00	1224930	99.6	99.6	99.6
	1.00	4649	.4	.4	100.0
	Total	1229579	100.0	100.0	

By combining the ILI designated illness with the flu, and sore throat admissions the following results are concluded:

THE_FLU

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	.00	1212224	98.6	98.6	98.6
	1.00	17355	1.4	1.4	100.0
	Total	1229579	100.0	100.0	

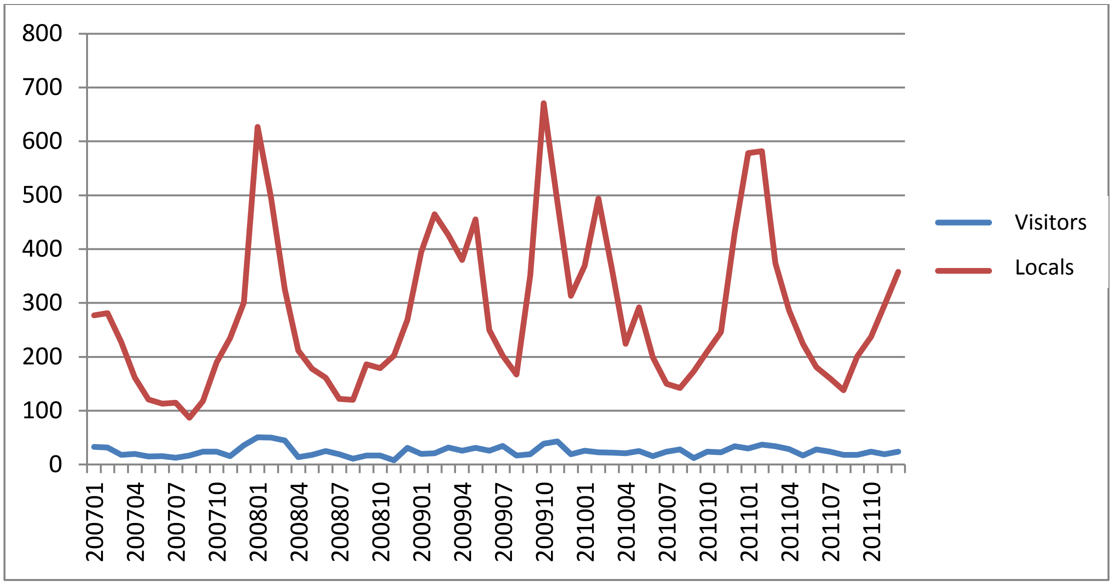
The following table shows a preliminary assessment of the cases classified as influenza for both visitors and local residents.

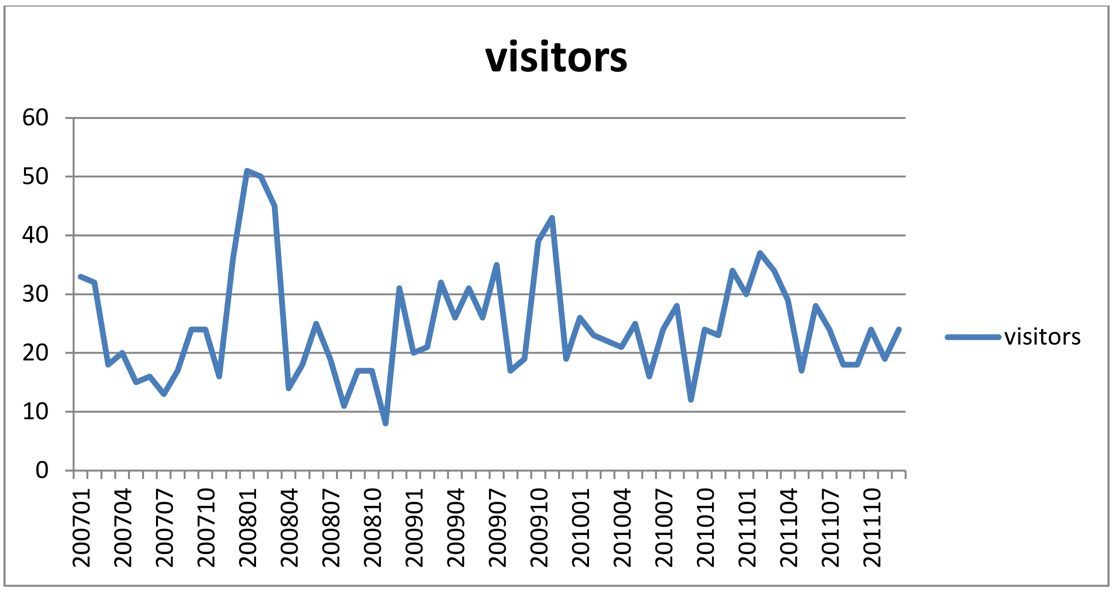
THE_FLU * local for Locals and Visitors

			local		Total
			0	1	
THE_FLU	Local	Count	178864	1033360	1212224
		% within THE_FLU	14.8%	85.2%	100.0%
	Visitors	Count	2011	15344	17355
		% within THE_FLU	11.6%	88.4%	100.0%
Total		Count	180875	1048704	1229579
		% within THE_FLU	14.7%	85.3%	100.0%

Flu Trends 2006-2010

In the two line graphs below, the trends for the outbreak of flu are illustrated. The first graph describes the frequency of flu tracking the outbreak between visitors and local residents. The next graph illustrates the trends for visitors based to provide a better relationship with the local resident trends. The graphs illustrate the changing basis of flu on an annual basis. In most years, outbreak among visitors peaked before the outbreak among local residents. However, during certain years, outbreaks among visitors seem to show a more erratic trend. This may be due to the time of year when certain outbreaks happen in different parts of the country. Further assessment of this data is warranted.





Limitations of the Data

There can be several important limitations to the data collected thus far. First, the data sets are large and many records require additional data cleansing to format data file mergers into a more reliable file. Because of the size of the data files, it is much more difficult to create accurate coding techniques to adequately capture chief complaints that might be indicated such as “flu”. For example, on examining all records related to “flu”, about 15% of the cases had to be omitted because of the inclusion of “fluid” or “flutter” in the chief complaint. Moreover, some terms such as “I feel terrible” might ultimately be coded as flu, but these are not captured in recoding string data into nominal data elements.

Second, any system based on hospital or clinic data has inherent delays based on the medical seeking behavior of the infected individual. In addition to the incubation period of the disease, there are delays in the seeking of medical care. The first step in a person's illness usually involves self-care and possibly over the counter (OTC) medications. This step may last from several hours to several days, and in many cases, is the only step involved in the infected person's medical care.

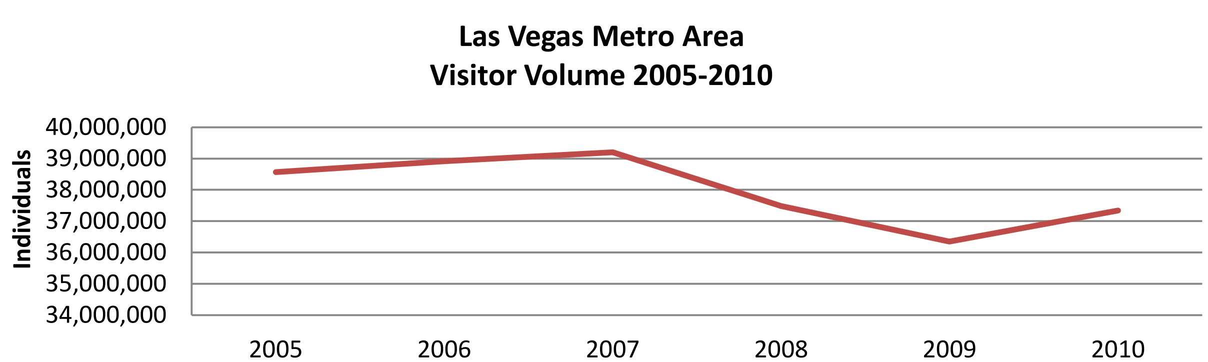
Third, if a person does decide to seek medical care, there are delays in transportation to the medical clinic and delays in the admissions process. These delays are usually not significant in the overall course of the illness, but are relevant to the frequency of data transmission and analysis. If data provided need to first be coded by hospital staff (such as an ICD9-CM diagnosis code), there are additional delays of hours to days.

Fourth, reliability of data - Some of the challenges to achieving real-time data surveillance when gathering information from EDs are that symptoms and CC are often recorded free-hand and there are no standardized terms so aggregating the data can become difficult. This is consistent with previous research regarding surveillance issues (Travers et. al, 2006). We also found that some information may take days or weeks to be transmitted due to not updating the patient record or deciding ICD-9 codes. Final diagnosis may depend on the reimbursement rates or how well the illness was charted. Although ICD-9 codes are standardized, the process of assigning patients ICD-9 codes involves multiple people and can take longer than desirable (Travers et. al, 2006).

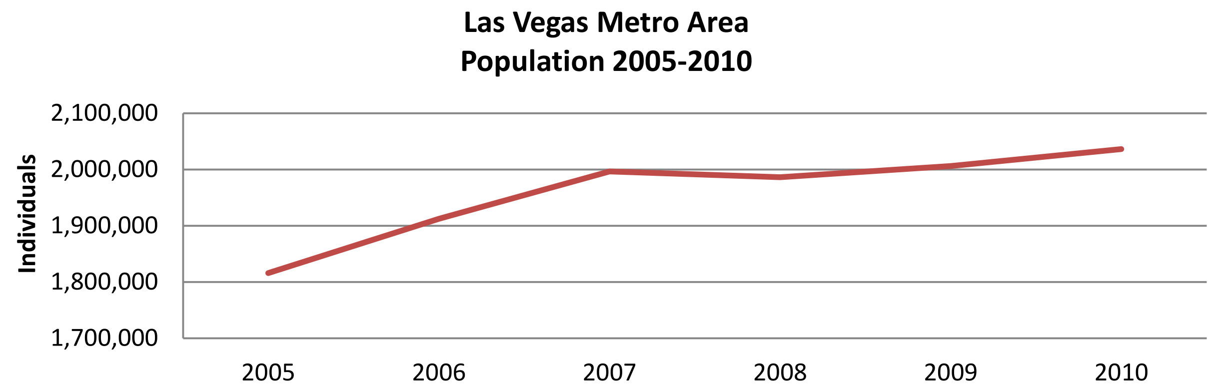
Much more work remains to be done on this study. The project team will delve further into the chief complaint data to make sure that we are able to identify more cases of flu or ILI that may be lost to data manipulation ore missing data fields. In addition, the team hopes to add additional missing data from the hospitals to make a more accurate time line calculation.

2.4.2 Travel and Disease Transmission

Based on the time frame of the sample provider data the simulator was staged with data representing resident, pass-through, and visitor travel for calendar years 2005-2010. An overview of the resident and visitor population change is provided in Exhibit 1.



	2005	2006	2007	2008	2009	2010
Series2	38,566,717	38,914,889	39,196,761	37,481,552	36,351,469	37,335,436



	2005	2006	2007	2008	2009	2010
Series2	1,815,700	1,912,654	1,996,542	1,986,146	2,006,347	2,036,358

	2005	2006	2007	2008	2009	2010
Annual population change %	3.90%	5.30%	4.40%	-0.50%	1.00%	1.50%
Annual population trend	68,675	96,954	83,888	-10,396	20,201	30,011
Avg. new residents per month	5,723	8,080	6,991	-866	1,683	2,501

Sources: GLS Research, U.S. Census Bureau, Nevada State Demographer, Clark County Comprehensive Planning, Las Vegas Convention and Visitors Authority.

Exhibit 1, Las Vegas Resident and Visitor Population 2005-2010

Demographic Overview

The predictive ILM simulates infectious disease status for individuals departing Las Vegas, and processes their travel route, mode of transportation, and destination. This data predicts the routes and paths of spread based on ground and air transportation bandwidth, demographics, traffic and airline data. Las Vegas receives almost 40 million visitors per year. That equates to approximately 100,000 visitors arriving and departing per day. GLS Research claims an average stay of approximately 3.5 days meaning there are typically about 300,000-350,000 visitors in Las Vegas at any given point.

The simulator produces visitor infection status and their mode, route, and time of departure. The output data allows analysis of the cities receiving exposed travelers including when they returned home. Exhibit 2 shows the top 32 cities receiving exposed from an outbreak simulation in Las Vegas based on the seasonal flu outbreak of 2008-2009

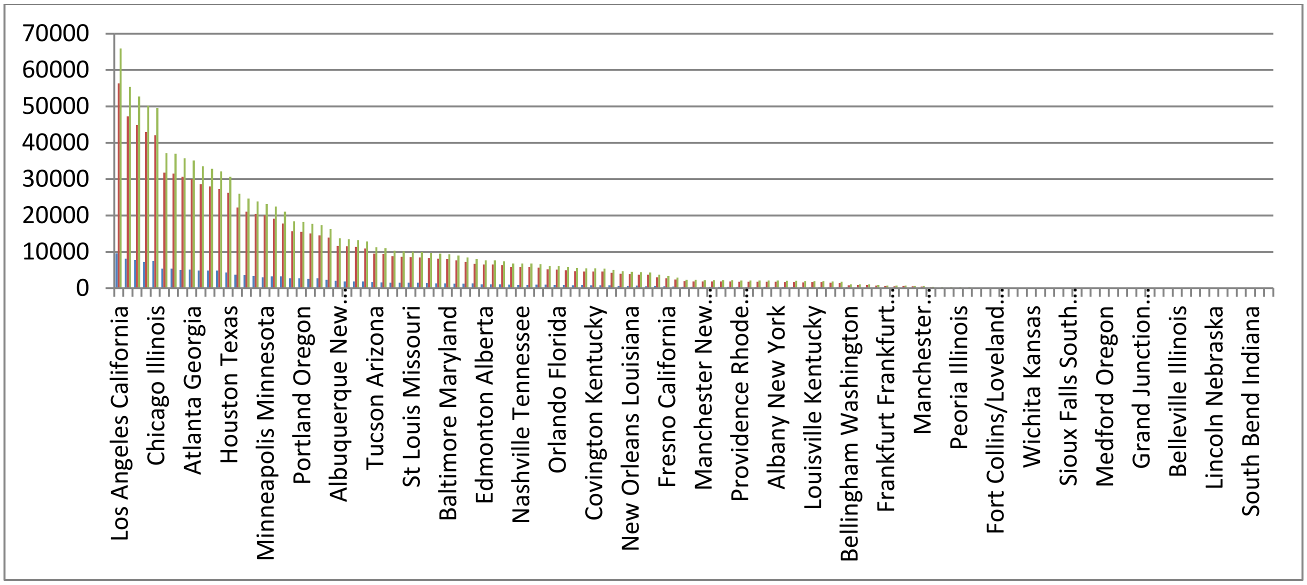


Exhibit 2, Top Destination for Exposed Individuals Departing Las Vegas by Air 2008-2009

Differences between the transportation total bandwidth and the visitors departing exposed are created by a stochastic simulation of interaction and effective contacts resulting in transmission, and the visitor's infection status and expected duration of infectivity. This allows modeling of heterogeneity for contacts, infectivity, and susceptibility. Based on this ILM approach and the stochastic infection simulation a list of cities receiving the most exposed visitors will not necessarily match a list of cities receiving the most passengers. Exhibit 3, *Summary of Top 50 Cities Receiving Simulated Infectious from Las Vegas by Air 2008-2009 Flu Season*, shows a difference in top cities from Exhibit 2 which shows the top 32 cities receiving exposed from an outbreak simulation in Las Vegas based on the seasonal flu outbreak of 2008-2009.

	Exposed	Infectious	Total
Los Angeles California	9654	56283	65937
Phoenix Arizona	8088	47247	55335
San Francisco California	7773	44909	52682
Denver Colorado	7244	42942	50186
Chicago Illinois	7487	42037	49524
Salt Lake City Utah	5375	31755	37130
San Diego California	5436	31562	36998
Dallas-Fort Worth Texas	5084	30662	35746
Atlanta Georgia	5116	30039	35155
New York New York	4899	28656	33555
Burbank California	4855	27984	32839
Seattle Washington	4840	27333	32173
Houston Texas	4384	26285	30669
Santa Ana California	3746	22200	25946
Reno Nevada	3634	21020	24654

San Jose California	3416	20482	23898
Minneapolis Minnesota	3035	20123	23158
Sacramento California	3315	19134	22449
Oakland California	3261	17821	21082
Ontario California	2756	15674	18430
Portland Oregon	2793	15481	18274
Philadelphia Pennsylvania	2589	15113	17702
Detroit Michigan	2799	14573	17372
Newark New Jersey	2363	13904	16267
Albuquerque New Mexico	2090	11634	13724
Charlotte North Carolina	1899	11591	13490
Vancouver British Colombia	1846	11410	13256
Toronto Ontario	1920	10939	12859
Tucson Arizona	1714	9556	11270
Washington District of Columbia	1585	9441	11026
Calgary Alberta	1529	8823	10352
Cleveland Ohio	1518	8625	10143
St Louis Missouri	1501	8585	10086
Kansas City Missouri	1539	8450	9989
Pittsburgh Pennsylvania	1430	8303	9733
Honolulu Hawaii	1401	8119	9520
Baltimore Maryland	1374	8026	9400
San Antonio Texas	1298	7696	8994
Indianapolis Indiana	1259	7258	8517
London West Sussex	1317	6763	8080
Edmonton Alberta	1127	6580	7707
Boston Massachusetts	1100	6589	7689
Milwaukee Wisconsin	1123	6328	7451
Austin Texas	988	5844	6832
Nashville Tennessee	959	5856	6815
El Paso Texas	951	5830	6781
Miami Florida	970	5664	6634
Columbus Ohio	964	5186	6150
Orlando Florida	946	5176	6122
Tampa Florida	943	4936	5879

Exhibit 3, Summary of Top 50 Cities Receiving Simulated Infectious from Las Vegas by Air 2008-2009 Flu Season

While the top cities receiving exposed returning Las Vegas visitors can be expected to receive thousands of exposed, many cities also receive exposed individuals. Exhibit 4, *Cities Receiving less than 1,000 Exposed 2008-2009 Simulation* lists some international and CONUS cities receiving exposed.

		Exposed	Infectious	Total
0809IMID	Incheon City Joong-Gu	142	750	892
0809IMID	Frankfurt Frankfurt Main	167	613	780
0809IMID	Santa Barbara California	192	568	760
0809IMID	Winnipeg Manitoba	123	612	735
0809IMID	Victoria British Colombia	153	542	695
0809IMID	Manchester Manchester	178	515	693
0809IMID	Regina Saskatchewan	88	269	357
0809IMID	Los Cabos San Jose del Cabo	103	234	337
0809IMID	Saskatoon Saskatchewan	90	229	319
0809IMID	Peoria Illinois	118	156	274
0809IMID	Kelowna British Colombia	43	221	264
0809IMID	Colorado Springs Colorado	109	145	254
0809IMID	Cedar Rapids Iowa	102	136	238
0809IMID	Fort Collins/Loveland Colorado	85	152	237
0809IMID	Springfield Missouri	87	144	231
0809IMID	Mc Allen Texas	85	141	226
0809IMID	Des Moines Iowa	88	135	223
0809IMID	Wichita Kansas	81	138	219
0809IMID	Stockton California	76	115	191
0809IMID	Missoula Montana	70	111	181
0809IMID	Santa Maria California	61	104	165
0809IMID	Sioux Falls South Dakota	56	106	162
0809IMID	Anchorage Alaska	27	103	130
0809IMID	Shreveport Louisiana	55	68	123
0809IMID	Great Falls Montana	58	62	120
0809IMID	Medford Oregon	55	63	118
0809IMID	Rochester Minnesota	52	66	118
0809IMID	Rapid City South Dakota	53	62	115
0809IMID	Redmond Oregon	44	68	112
0809IMID	Grand Junction Colorado	39	70	109
0809IMID	Hermosillo Sonora	33	75	108
0809IMID	Idaho Falls Idaho	45	63	108
0809IMID	Laredo Texas	50	57	107
0809IMID	Belleville Illinois	46	57	103
0809IMID	Fargo North Dakota	47	56	103
0809IMID	Chicago/Rockford Illinois	48	54	102
0809IMID	Pasco Washington	50	52	102
0809IMID	Lincoln Nebraska	51	47	98
0809IMID	Green Bay Wisconsin	41	55	96
0809IMID	Bismarck North Dakota	35	59	94

0809IMID	Duluth Minnesota	35	56	91
0809IMID	South Bend Indiana	35	50	85
0809IMID	Eugene Oregon	36	44	80
0809IMID	Billings Montana	40	32	72

Exhibit 4, Cities Receiving less than 1,000 Exposed 2008-2009 Simulation

Simulation air versus road visitors and the main paths of egress are compared in Exhibit 5, *Exposed Road Traveler Routes 2005-2006 Simulation* and Exhibit 6, *Exposed Air Traveler Destinations 2005-2006 Simulation*. According to GLS Research annual demographic reports approximately 54 % of visitors travel by ground transportation and 46% by air.

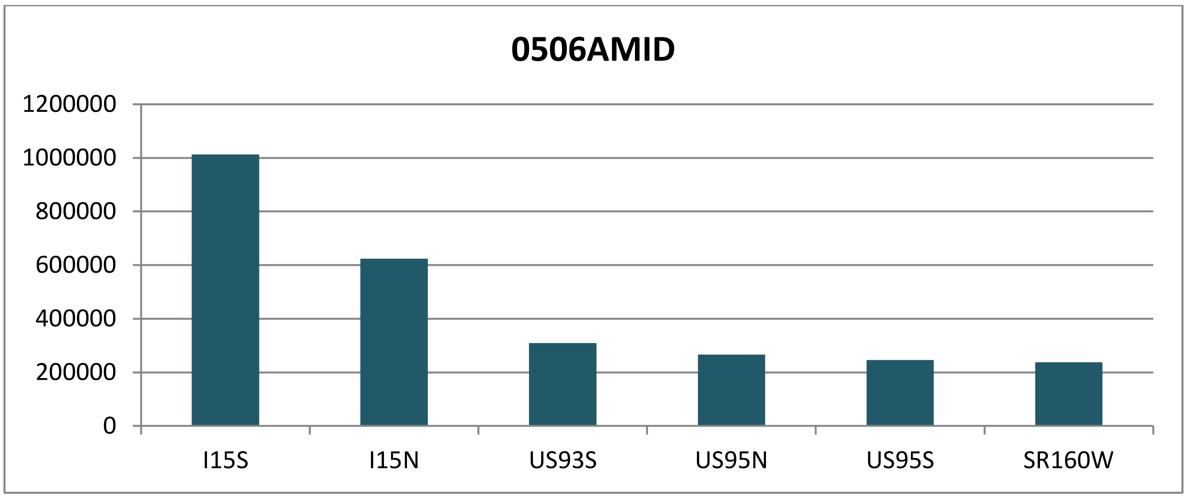


Exhibit 5, Exposed Road Traveler Routes 2005-2006 Simulation

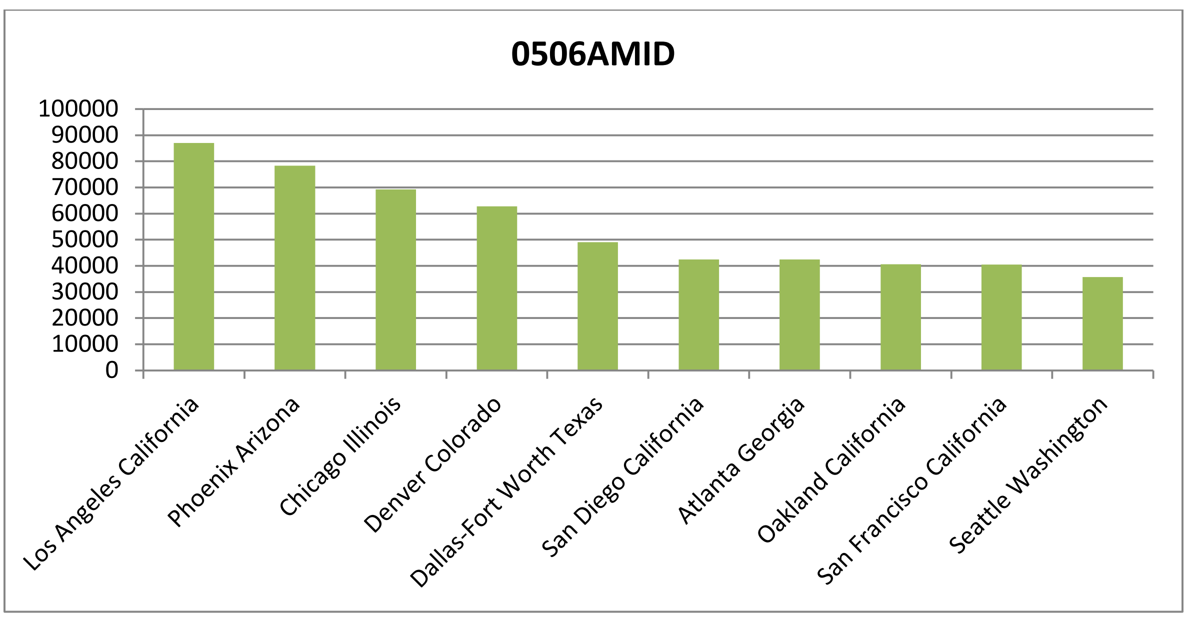
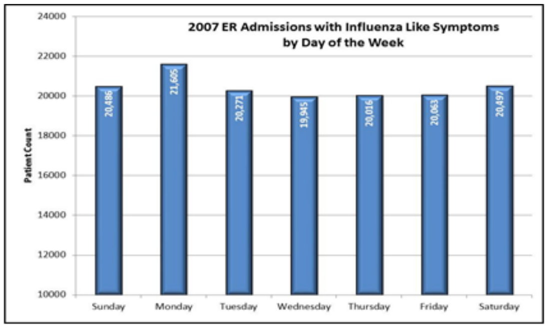
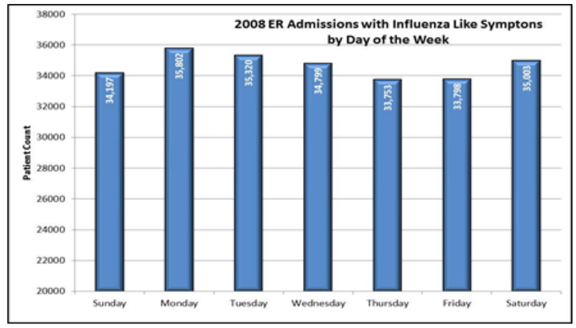


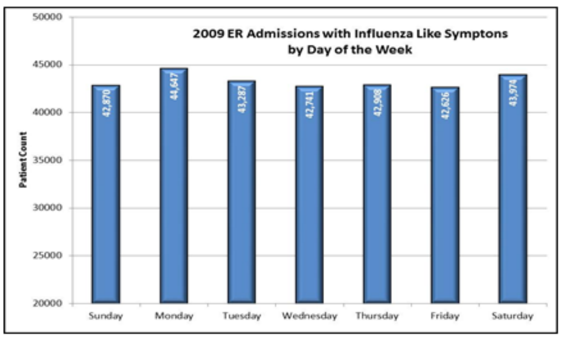
Exhibit 6, Exposed Air Traveler Destinations 2005-2006 Simulation

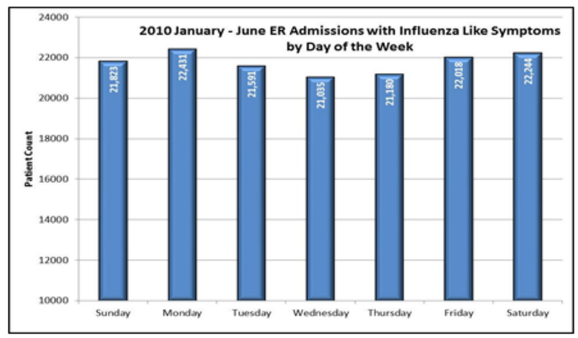
Effects within the Sample

Day of the week and holiday effects are present in the sample. Burkom's (2006) Monday spike is visible as are additional noise components. Exhibits 7 and 8 summarize the DOW and Holiday effects on case reports classified ILI by the study.









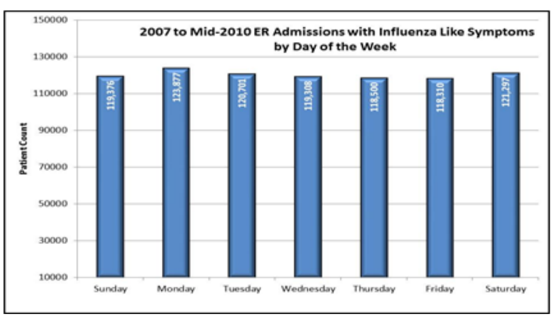


Exhibit 7, Day of Week Effects within the Sample

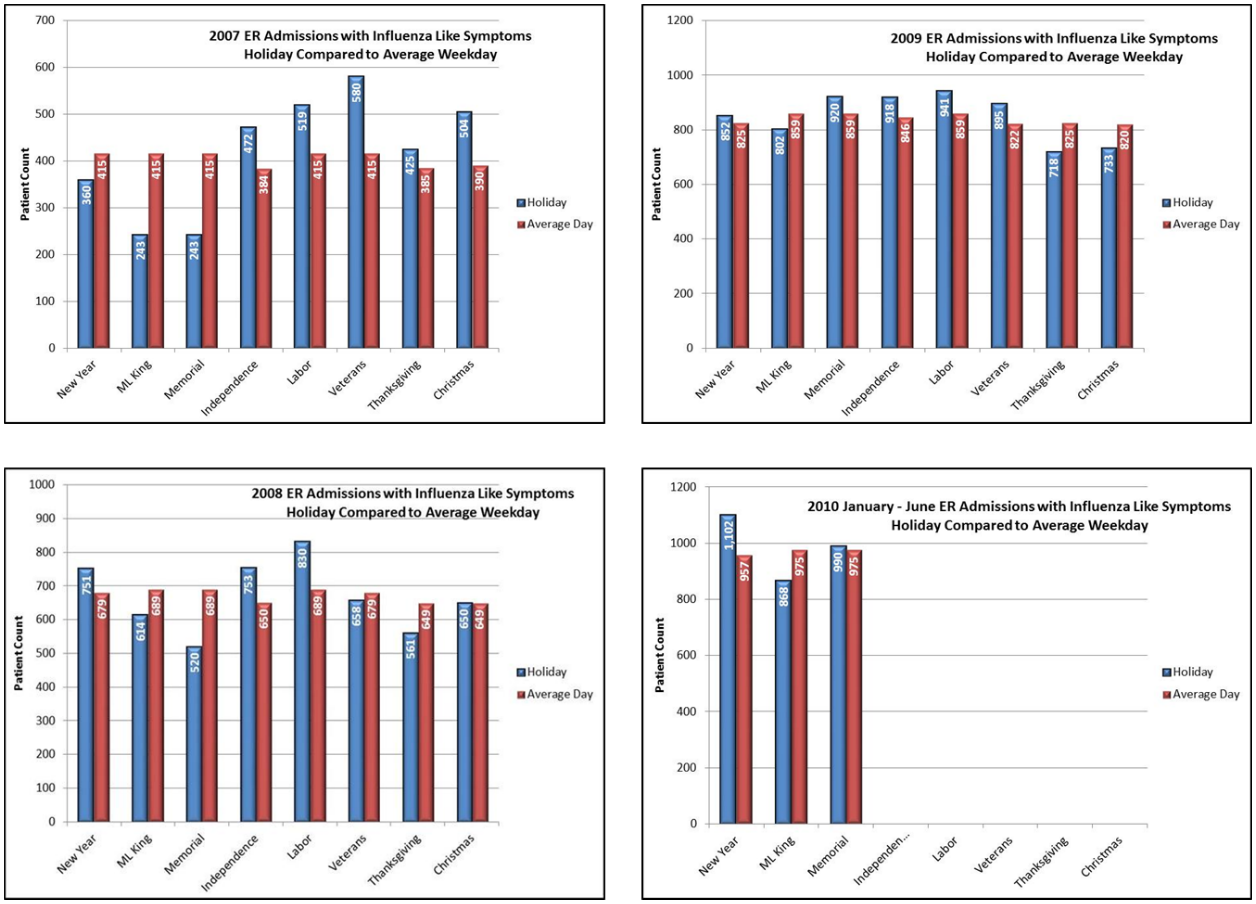


Exhibit 8, Holiday Effects within the Sample by Year

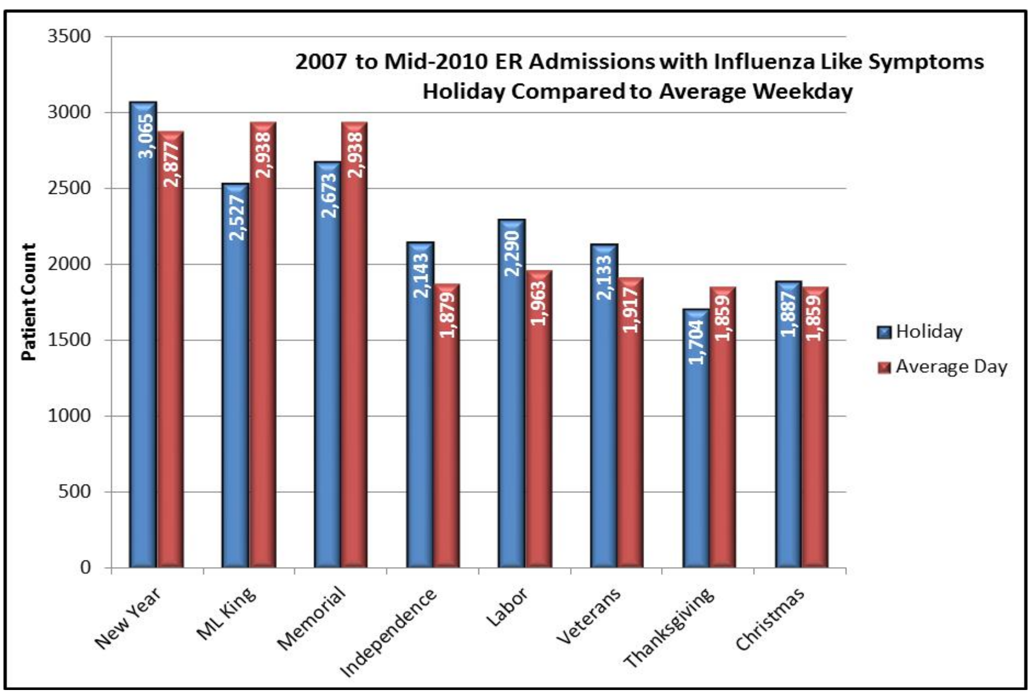


Exhibit 9, Holiday Effects within the Sample 2007-mid2010

3.0 Key Research Accomplishments

Completed investigative meetings with hospitals, clinics, physician practices, paramedics, Nevada Department of Transportation, airport, and hospitality industry representatives

Updated surface travel database and added second half 2008 and all 2009 and 2010 information

Updated air travel database for system test adding 2009 and 2010 data

Updated simulator ILI files using CDC sentinel data for 2008, 2009, and 2010

Prepared and maintained message server, provider (UMC) ED data interface, and database

Continued requirements analysis and updated system functional requirements

Conducted and documented a literature review of related research and publications

Conducted an empirical study of Las Vegas Strip employment including non-resort business, convention attendance, and interaction between residents and visitors to improve understanding of contact rates

All staff completed two CITI training courses for research protection

Updated and submitted protocol to UNLV IRB for approval to access and use provider ED data

Received UNLV IRB protocol approval

Submitted UNLV IRB approved protocol to Office of Research Protection for approval to access and use provider ED data

Received ORP decision of Non-Human Use data

Completed ED data normalization, anomaly removal, binning of syndromes, and preliminary data analyses in preparation for test

Evaluated some available, existing biosurveillance codes for suitability including SYDOVAT, Trisano, Real-time Outbreak Detection System, EpiFire, Global Epidemic Model and Global Influenza Surveillance Network

Ported and tested synthetic data generation codes using R to prepare synthetic test data sets with appropriate distributions and effects

Ported MATLAB MCUSUM and MEWMA codes to Octave

Developed EWMA and CUSUM detection codes

Developed software code for state-space disease model with mobility between cities and models for SECIR adding carrier-latency and SEInR including variable infectivity

Modified software codes for simulation of air and road travel to improve performance. Converted single-computer designed codes to run on the Hadoop cluster for performance improvement and developed some of the new modules required to run biostage codes on the cluster

Updated the Hadoop cluster hardware to reduce travel simulation time

4.0 Reportable Outcomes

Received Non-Human Use ruling from Office of Research Protection

Established interface with the County hospital system and obtained and stored year of ED data

Obtained ED data from University Medical Center, Sunrise hospital, and three Valley Health Systems hospitals

Completed prototype software for modeling population mobility and correlation of travel and outbreak information sets

Prepared test software codes for outbreak detection and conducted initial validation testing

Completed prototype software for modeling population mobility and simulating outbreaks

5.0 Conclusions

Meaningful integration of travel and infectious disease propagation information is highly applicable to effective epidemiology. The development and integration of surveillance with population dynamics, especially travel, should be considered essential function for effective epidemiology in the computer age.

Data shaping costs along with legitimate privacy concerns and the lack of mandated standards of reporting and recordkeeping result in surveillance-functions receiving a very poor signal in a very noisy environment. The main factors limiting progress are legislative, but technical advancements are also needed.

Individual-level models (ILM) enable modeling of heterogeneity and statistical distance not possible in meta-population infection spread models. Advancements in data processing technology enable and therefore mandate development of improved data processing methods and new infection models. The resource requirements for ILM modeling are no longer a constraint but efficient, validated methods for data integration and shaping are a required, complementary component.

Regional human daily population variance is a significant noise component within syndromic time series. This effect has potential within the research domain for filtering or providing explanation and within the surveillance function to expand situational awareness capacity.

Research in these areas is essential and should continue.

The results of this study have not been validated. Tests were ongoing in parallel with the development of this report. Data access was delayed much longer than scheduled awaiting an ORP review. This compressed the schedule. The ORP review determined the sample was non-human-use.

This report was concluded based on the expiration of resources for the level-of-effort.

6.0 References

Anoli, Luigi; Duncan, Starkey Jr.; et al. *The Hidden Structure of Interaction.* IOS press. Amsterdam, Berlin, Oxford, Tokyo, Washington, D. C. 2005.

Armbruster, Benjamin and Bandeau, Margaret L. “Contact tracing to control infectious disease: when enough is enough”. *Health Care Management Science* 10 (October 2007): 341-355. Available online: <http://www.stanford.edu/dept/MSandE/cgi-bin/people/faculty/brandeau/pdfs/Armbruster%20HCMS%20article.pdf>.

Atkinson, Michael P. and Wein, Lawrence M. “Quantifying the Routes of Transmission for Pandemic Influenza”. *Bulletin of Mathematical Biology* 70 (2008): 820-867.

Barker J, Vipond IB, Bloomfield SF. “Effects of cleaning and disinfection in reducing the spread of norovirus contamination via environmental surfaces”. *J Hosp Infect.* 2004; 58:42–9.

Biosurveillance Data Steering Group (BDSG). Minimum Data Set. Available online: <http://www.docstoc.com>.

Boëlle PY, Bernillon P, Desenclos JC. “A preliminary estimation of the reproduction ratio for new influenza A (H1N1) from the outbreak in Mexico, March-April 2009”. *Euro Surveill.* 2009;14(19):pii=19205. Available online: <http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=19205>.

Bravata, D., McDonald, K., et al (2004) “Systematic Review: Surveillance Systems for Early Detection of Bioterrorism-Related Diseases”. *Ann Internal Medicine* 140.910-922. American College of Physicians.

Brooks, Jennifer. Air Transport Association, “Control of Communicable Diseases (Q Rule). Air Transport Association.

Brownstein, J.S., C.J. Wolfe, and K.D. Mandl (2006). “Empirical evidence for the effect of airline travel on inter-regional influenza spread in the United States”. *PLOS Medicine*; 3(10): Accessed July 26, 2008.from <http://medicine.plosjournals.org/perlserv/>.

Burkom, H.S. and Shmueli, Galit. "Statistical Challenges Facing Early Outbreak ... Surveillance Conference (in Advances in Disease Surveillance, 1:53), 2006.

Busenberg, Stavros and van den Driessche, P. “Nonexistence of Periodic Solutions for a Class of Epiedemological Models.” Available at: dspace.library.uvic.ca:8080/bitstream/handle/.../DMS%20539.pdf?...

Cambridge Systematics, Inc. (2010), “Travel Model Validation and Reasonableness Checking Manual”. Available online: <http://tmip.fhwa.dot.gov/resources/clearinghouse/docs/FHWA-HEP-10-042/FHWA-HEP-10-042.pdf>.

Cameron, Wendy K. “Public Health Planning for Vulnerable Populations and Pandemic Influenza”. Master’s Thesis. Naval Postgraduate School. Monterey, Ca. 2008.

Camitz, Martin. “Computer Aided Infectious Disease Epidemiology – Bridging to Public Health”. Karolinska Institutet. Stockholm, 2010. Available online: <http://publications.ki.se/jspui/bitstream/10616/40323/1/Thesis%20Camitz>.

Carley, K., Fidsma, D., et al. (2004) “BioWar: Scalable Multi-Agent Social and Epidemiological Simulation of Bioterrorism Events”. Electronic Publication, Pittsburgh, PA, IEEE SMCA03-11-0274. Centers for Disease Control and Prevention: Seasonal Influenza (Flu). Flu Activity and Surveillance. Accessed June 25, 2008 from: <http://www.cdc.gov/flu/weekly/fluactivity.htm.>.

Cassa, A. Christopher “Spatial Outbreak Detection Analysis Tool: A system to create sets of semi-synthetic geo-spatial clusters.” (August 30, 2004) Available online: <http://groups.csail.mit.edu/medg/ftp/cassa/meng.pdf>.

Ceccine, Gary & Moore Melinda. “Infectious Disease and National Security; Strategic Information Needs”. Rand National Defense Research Institute, 2006. Accessed April 18, 2011 from: <http: //www.rand.org>.

Centers for Disease Control and Prevention: “Seasonal Flu. Flu symptoms and activity”. Accessed April 18, 2008 from: <http://www.cdc.gov/flu/weekly/fluactivity.htm>.

Centers for Disease Control and Prevention (2009). “H1N1 Flu”. Accessed November 12 2008 from <http://www.cdc.gov/h1n1flu/qa.htm>.

Chang, Hwa-Gan PhD, Jian-Hua Chen MD, et al. “The Availability of Data Elements in an Existing Emergency Department Electronic Medical Record System”. New York State Department of Health, Albany, NY and Emergency Medical Associates of New Jersey Research Foundation, New Jersey

Chowell, Gerardo; Hyman, James M.; and Bettencourt, Luis M.A. “Mathematical and Statistical Estimation Approaches in Epidemiology” Dordrecht Heidelberg London New York: Springer, 2009. Available online: <http://books.google.com/books?id=DYSLbyq_hYgC&pg=PA124&dq=Contact+rate+infectious+disease&hl=en&ei=Szv6TftaxPDSAevDkL0D&sa=X&oi=book_result&ct=result&resnum=3&ved=0CDkQ6AEwAjge#v=onepage&q=Contact%20rate%20infectious%20disease&f=false>.

Conway, Mike; Dowling, John; and Chapman, Wendy. “Developing a Biosurveillance Application Ontology for Influenza- Like- Illness.” *Proceedings of the 6th Workshop on Ontologies and Lexical Resources* (Ontolex 2010), pp. 58–66. Beijing, August 2010

Cooper B, Pitman R, Edmunds W, & Gay N. (2006) Delaying the International Spread of Pandemic Influenza: *PLoS Medicine* vol. 3 no. 6:e212 pp. 845-0854.

Cooper, D.L.; Verlander, N. Q.; Elliot, A. J.; Joseph, C. A.; and Smith, G. E. “Can syndromic thresholds provide early warning of national influenza outbreaks?” Journal of Public Health, Volume 31, Issue 1, pp17-25. Available online: <http://jpubhealth.oxfordjournals.org/content/31/1/17.full>.

Cowgill, K., Lucas, C., Benson, R., et al. (2005). “Recurrence of legionnaire’s disease at a hotel in the United States Virgin Islands over a 20-year period”. *Clinical Infectious Disease.* 2005:40 pp. 1205-1207.

Cowling, Benjamin J.; Wong, Irene O. et al “Methods for monitoring influenza surveillance data”. *Oxford Journals International Journal of Epidemiology*. 35 (July 2006) 1314-1321. Available online: <http://ije.oxfordjournals.org/content/35/5/1314.full>.

Crepey, P. and M. Barthelemy (2007). “Detecting robust patterns in the spread of epidemics: a case study of influenza in the United States and France”. *American Journal of Epidemiology*, 166(11): 1244-51.

“Data Sources: County Data – US Census Bureau American Community Survey”. Available online: <http://factfinder.census.gov/home/saff/main.html?_lang=en&_ts=>.

Delaney, John B., Jr. “The National Disaster Medical Systems Reliance on Civilian-Based Medical Response Teams in a Pandemic is Unsound”. Master’s Thesis. Naval Postgraduate School. Monterey Ca. 2007.

Department of the Army. Army Profile FY05. Retrieved November 14, 2009 from <http://www.armyg1.army.mil/hr/docs/demographics/FY05%20Army%20Profile.pdf>.

Diekmann, O. “Mathematical Epidemiology of Infectious Disease”. *Images of SMC Research*. 1996.

Dippold, L., Lee, R., Selman, C., Monroe, S., & Henry, C. (2003). “A gastroenteritis outbreak due to norovirus associated with a Colorado hotel”. *Journal of Environmental Health.* 66(5): 13-7.

Elveback, Lila R. “Simulation of Stochastic Discrete-Time Epidemic Models for Two Agents.” *Advances in Applied Probability* Vol 3, Number 2 (Autumn 1971): pp 226-228.

Epstein, J.M. D.M. Goedecke, Yu, et al, (2007). "Controlling pandemic flu: the value of international air travel restrictions". *PLoS One*, 5(e141). Accessed July 25, 2008 from: <www.plosne.org>.

Flahault, A.; Vergue, E. et al (2009). "Potential for a global dynamic of Influenza A (H1N1)". *BMC Infectious Diseases:* 9:129.

Foley, John R. "The Pandemic Pendulum: A Critical Analysis of Federal and State Preparedness for a Pandemic Even". Masters thesis. Naval Postgraduate School. Monterey, Ca. 2009.

Fraser, Christopher; Riley, Steven; et al. "Factors that make an Infectious Disease Outbreak Controllable". *Proceedings of the National Academy of Sciences of the United States of America* 101 (16) (April 2004) 6146-6151. Available online: <www.ncbi.nlm.nih.gov/pmc/articles/PMC395937/>.

Fricker, Ronald D., Jr. "Some Methodological Issues in Biosurveillance". Naval Postgraduate School Monterey, CA. 2009

Grais, R.F, J.H. Ellis, A. Kress, G.E. Glass (2004). "Modeling the spread of annual influenza epidemics in the U.S.: the potential role of air travel". *Health Care Management Science,* 7(2): 127-34.

Haber, Michael J.; Shay, David K.; et al. "Effectiveness of Interventions to Reduce Contact Rates during a Simulated Influenza Epidemic". *Emerging Infectious Diseases* (Volume 13, No.4) April 2007. Available online: <http://www.cdc.gov/EID/content/13/4/pdfs/581.pdf>.

Hannon, Bruce and Matthias, Ruth. *Modeling Dynamic Biological Systems*. New York: Springer-Verlag 1999. Available online: <http://books.google.com/books?id=CCkNkjnE2DUC&pg=PA131&dq=Contact+rate+infectious+disease&hl=en&ei=XA_xTcj2PPSr0AHHluWlBA&sa=X&oi=book_result&ct=result&resnum=3&ved=0CD0Q6AEwAjgK#v=onepage&q=Contact%20rate%20infectious%20disease&f=false>.

Henderson, J. (2004). "Paradigm shifts: National tourism organizations and education and healthcare tourism. The case of Singapore". *Tourism and Hospitality Research* 5(2), 170-180. Retrieved November 14, 2009 from ABI/INFORM Global. (Document ID: 694775801).

Hethcote, Herbert W. "The Basic Epidemiology Models: Models, Expressions for *R*0, Parameter Estimation, and Applications." *Mathematical Understanding of Infectious Disease Dynamics*. World Scientific Publishing Co. Pte. Ltd. Available online: <http://www.worldscibooks.com/mathematics/7020.html>.

Hu, Cecilia X and Knitt, Matthew C. " A Comparative Analysis of Multivariate Statistical Detection Method." Naval Postgraduate School, Monterrey, CA, 2007. Available at: www.dtic.mil/cgi-bin/GetTRDoc?AD=ADA470074

Hufnagel L, Brockmann D, and Geisel T. (2004) "Forecast and control of epidemics in a globalized world". *PNAS,* vol. 101 no. 42:15124-15129.

Hyman, James M. and Li. Jai. "Differential Susceptibility and Infectivity Epidemic Models". *Mathematical Biosciences and Engineering,* Volume 3, Number 1 (January 2006): 89-100. Available online: <http://www.mbejournal.org/>.

Joner, M.D., Jr., Woodall, W.H., Reynolds, M.R. Jr., and R.D. Fricker, Jr. (2006). The Use of Multivariate Control Charts in Public Health Surveillance (draft).

Kontzer, T. (2004). "Privacy pressure". *Information Week:* 981; ABI/INFORM Global p. 22.

Korotayev, A.; Malkov, A.; and Khaltourina, D. *Introduction To Social Macrodynamics: Secular Circles and Millenial Trends.* URSS, Moscow 2006

Kulldorff M, Heffernan R, Hartman J, Assunção RM, Mostashari F. (2005). "Space-Time Permutation Model a space-time permutation scan statistic for the early detection of disease outbreaks". *PLoS Medicine,* 2:216- 224.

Longini, Ira M.; Halloran, M. Elizabeth, et al "Containing Pandemic Influenza with Antiviral Agents". *American Journal of Epidemiology* 159 (April 1, 2004): 623-633.

Lopez-Cervantes, M., Venado, A., Moreno, et al. (2009). "On the spread of the novel Influenza A (H1N1) virus in Mexico". *J Infect dev Ctries* 2009*;* 3 (5): 327-330.

Ma, Zhien and Li, Jia. *Dynamical Modeling and Analysis of Epidemics.* World Scientific Publishing. Hackensack, N.J. 2009.

Maciejewski, Ross; Hafen, Ryan; et al. (May/June 2009). "Generating Synthetic Syndromic-Surveillance Data for Evaluation Visual-Analytics Techniques", *IEEE Computer Society.*

Magnusson, Magnus S. "Understanding Social Interaction: Discovering Hidden Structure with Model and Algorithms." *The Hidden Structure of Interaction: From Neurons to Culuure Patterns.* IOS Press, Amsterdam 2005

Min, J. (2008). "Forecasting Japanese tourism demand in Taiwan using an intervention analysis". *International Journal of Culture, Tourism and Hospitality Research*, 2(3), 197-216. Retrieved November 14, 2009 from ABI/INFORM Global. (Document ID: 1550254771).

MMWR (2007). Retrieved November 13, 2009 from <http://www.cdc.gov/mmwr/preview/mmwrhtml/mm5648a2.htm>.

Moser, M.R., T. R. Bender, H. S. Margolis, G.R. Noble, A.P. Kendal, D.G. Ritter (1979). “An outbreak of influenza aboard a commercial airliner”. *American Journal of Epidemiology*, Vol. 110 (1): 1-6.

Nelson, Kenrad E. and Williams Carolyn Masters. *Infectious Disease Epidemiology*. Sudbury MA: Jones and Bartlett Publishers, 2007. Available online: <http://books.google.com/books?id=o_j-G4zJ4cQC&pg=PA198&lpg=PA198&dq=Contact+rate+infectious+disease&source=bl&ots=RFylSiYDxZ&sig=MfbdapJ2JS_7CeS-OeYZ29bqYqU&hl=en&ei=n0juTYzQNYfJgQeh47GVDw&sa=X&oi=book_result&ct=result&resnum=2&ved=0CCgQ6AEwATgK#v=onepage&q=Contact%20rate%20infectious%20disease&f=false>.

Okhmatovskaia, A., Chapman, W., Collier, N.,Espino, J., and Buckeridge, D. (2009). “SSO: The Syndromic Surveillance Ontology.” *Proceedings of the International Society for Disease Surveillance.*

Pelecanos, Anita M.; Ryan, Peter A.; Gatton, Michelle L. “Outbreak detection algorithms for seasonal disease data: a case study using ross river virus disease”. Available online: <http://www.biomedcentral.com/1472-6947/10/74>.

Pine, R. & McKercher, B. (2004). “The impact of SARS on Hong Kong’s tourism industry”. *International Journal of Contemporary Hospitality Management.* Vol 16. Issues 2 139-143.

Rashid, H. S. Shafi, R. Booy, H.E. Bashir, et al., (2008). Influenza and respiratory syncytial virus infections in British hajj pilgrims. Emerging Health Threats Journal, 1(e2). Accessed July 25, 2008 from: <www.eht-journal.org>.

Reed C, Angulo FJ, Swerdlow DL, et al. (2009). “Estimates of the prevalence of pandemic (H1N1) 2009, United States, April–July 2009”. *Emerg Infect Dis* [serial on the Internet]. 2009 Dec. Retrieved November 14, 2009 from: <http://www.cdc.gov/eid/content/15/12/pdfs/09-1413.pdf>.

Ritzwoller, Debra P.; Kleinman, K.; et al. “Comparison of Syndromic Surveillance and a Sentinel Provider in Detecting an Influenza Outbreak --- Denver, Colorado, 2003.”*MMWR Supplement 54(suppl)* (August 26, 2005): 151-156. Available online: <http://www.cdc.gov/Mmwr/preview/mmwrhtml/su5401a24.htm>.

Roberts, Anthony. “The Emergence of Disease in Early World-Systems: a Theoretical Model of World-System and Pathogen Evolution”. Presentation at the annual meetings of the

American Sociological Association. Atlanta, Ga. August 2010. Available online: <http://irows.ucr.edu/papers/irows62/irows62.htm>.

Joseph, C., Ricketts, K. (2007). “From development to success: the European surveillance scheme for travel associated Legionnaires' disease”. *European Journal of Public Health*, 17(6), 652-6. Retrieved November 15, 2009 from ABI/INFORM Global.

“Riviera Hotel & Casino Selects NFR Security to Protect Guest and Financial Data”. *Business Wire*. New York: Apr 5, 2004. p. 1

Rota MC, Caporali M, Massari M.,”European Guidelines for Control and Prevention of Travel Associated Legionnaires' Disease: the Italian experience”. *Euro Surveill.* 2004;9(2):pii=445. Avalable online: <http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=445>.

Rothman, Kenneth J.; Greenland, Sander; and Lash, Timothy L. *Modern Epidemiology*. Philadelphia: Lippincott Williams & Wilkins, 2008. Available online: <http://books.google.com/books?id=Z3vjT9ALxHUC&printsec=frontcover&dq=inauthor:%22Kenneth+J.+Rothman%22&hl=en&ei=mnb3TZH7CYT00gGywbDFCw&sa=X&oi=book_result&ct=result&resnum=1&ved=0CDAQ6AEwAA#v=onepage&q&f=false>.

Russell, C.A., T.C. Jones, I.G. Barr, et al. (2008). “The global circulation of seasonal influenza a (H2N2) viruses”. *Science*, 320; April 18, 2008: 340-46.

Rvachev, Leonid A. and Longini, Ira M. “AMathematical Model For The Spread of Influenza.” *Mathematical Biosciences* (75) Issue 1 July 1985: pp.3-22.

Sattenspiel, Lisa and Klaus Dietz (1995) “A structured epidemic model incorporating geographic mobility among regions”. *Mathematical Biosciences* 128:71-91.

Scholer, M. (2004). “Development of a Syndrome Definition for Influenza-Like-Illness.” *Proceedings of American Public Health Association Meeting* (APHA 2004).

Shmueli, Galit. "Wavelet-Based Monitoring in Modern Biosurveillance", Working Paper, RHS-06-002, Robert H Smith School, University of Maryland (http://ssrn.com/abstract=902878).

Smith, R. (2004). “Courts: FBI can’t force casinos to reveal records”. Retrieved on November 20, 2009 from: <http://www.casinocitytimes.com/news/article/courts-fbi-cant-force-casinos-to-reveal-records-145652>.

Southern Nevada Health District (2007). “Guidelines for the prevention and control of norovirus in hotels/casinos”. Retrieved November 15 from: <http://southernnevadahealthdistrict.org/health-topics/norovirus.php>.

“Surveillance Group for New Influenza A (H1N1) Virus Investigation and Control in Spain. New influenza A (H1N1) virus infections in Spain, April-May 2009”. *Euro Surveill*. 2009;14(19):pii=19209. Avalable online: <http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=19209>

Tourism Data: LVCVA – 2008 Las Vegas Visitor Profile Available online: <http://www.lvcva.com/getfile/2008%20Las%20Vegas%20Visitor%20Profile.pdf?fileID=107)>.

Travers, D., Barnett, C., Ising, A. & Waller, A. (2006). Timeliness of emergency department diagnosis for syndromic surveillance. *AMIA 2006 Symposium Proceedings,* 769-773.

Vidal, R. & Lawson, A. (2006). “Online updating of space-time disease surveillance models via particle filters”. *Stat Methods Med Res*.; 15(5):423-44.

Wagner, M., Moore, M. and Aryel, R. editors, (2006) *Handbook of Biosurveillance*, Elsevier Academic Press, Burlington, MA.

Wallstrom, Garrick, L., Wagner, M., Hogan, W., “High-Fidelity Injection Detectability Experiments: a Tool for Evaluating Syndromic Surveillance Systems”. RODS Laboratory, University of Pittsburgh, Pittsburgh, Pennsylvania. Available online: <http://www.cdc.gov/Mmwr/preview/mmwhtml/su5401a15.htm>.

Wenzel, R. P. (1996). “Airline travel and infection”. *New England Journal of Medicine*. 334(15):981-2.

World Health Organization (2009). Accessed November 5, 2009 from: <http://www.who.int/csr/don/2009_11_06/en/index.html>.

7.0 Project Personnel

Nicholas A. CerJanic, MS, Principal Investigator
QNA Director, Information Systems

Chris Cochran, Ph.D. UNLV Principal Investigator
Department of Health Care Administration and Policy

Paulo Pinheiro, Ph.D., UNLV Co-Principal Investigator
School of Community Health Sciences, Epi Program

Henry J. Osterhoudt, PhD
QNA Engineer/Historian
Public Outreach Lead

Geoffery Roberts, BA
QNA Software Engineer

John Goetz, MS
Consultant

Dominic Henriques, MHA
Research Assistant, UNLV Department of Health Care Administration and Policy

Ronald Ryan, MS
QNA Systems Engineer

Kelly Thomas, BS
QNA Software Tester

Shaun Galey, AAS
QNA Software Engineer

Appendix A: Letter Report from Colleagues

Directly from formal email correspondence dated 17 May 2012:

To: Nick CerJanic, Qinetiq-NA
From: Chris Cochran, Ph.D., Paulo Pinheiro, PhD and Dominic Henriques
Date: May 17, 2012
Subject: Analysis of ILI Outbreak for October 1, 2008 – September 30, 2009

In table 1, we show the number of cases of flu for September 28, 2008 through October 3, 2009. These dates represent a 52-week period to reflect the period requested with each week beginning on a Sunday. For comparison, we took a five year average of the number of cases to estimate outbreak starts. The trend for increases in the cases of ILI begins Dec. 28, 2008 and peaks the week of March 2, 2009. There is another spike on April 26, 2009 which drops off suddenly. The researchers believe that this spike is an aberration due to reports of the H1N1 virus that hit the news wires precisely at this time. It is also worth noting that each hospital submitting data showed a dramatic two-day increase in the number of visits for ILI during that period. In our estimation, this aberration was caused by the "worried well", since the cases drop off quickly and the first H1N1 cases were not reported in Nevada until later in the summer of 2009. However, there does appear to be another outbreak in late September 2009. The researchers believe that this outbreak is more closely related to the number of actual H1N1 cases during that year.

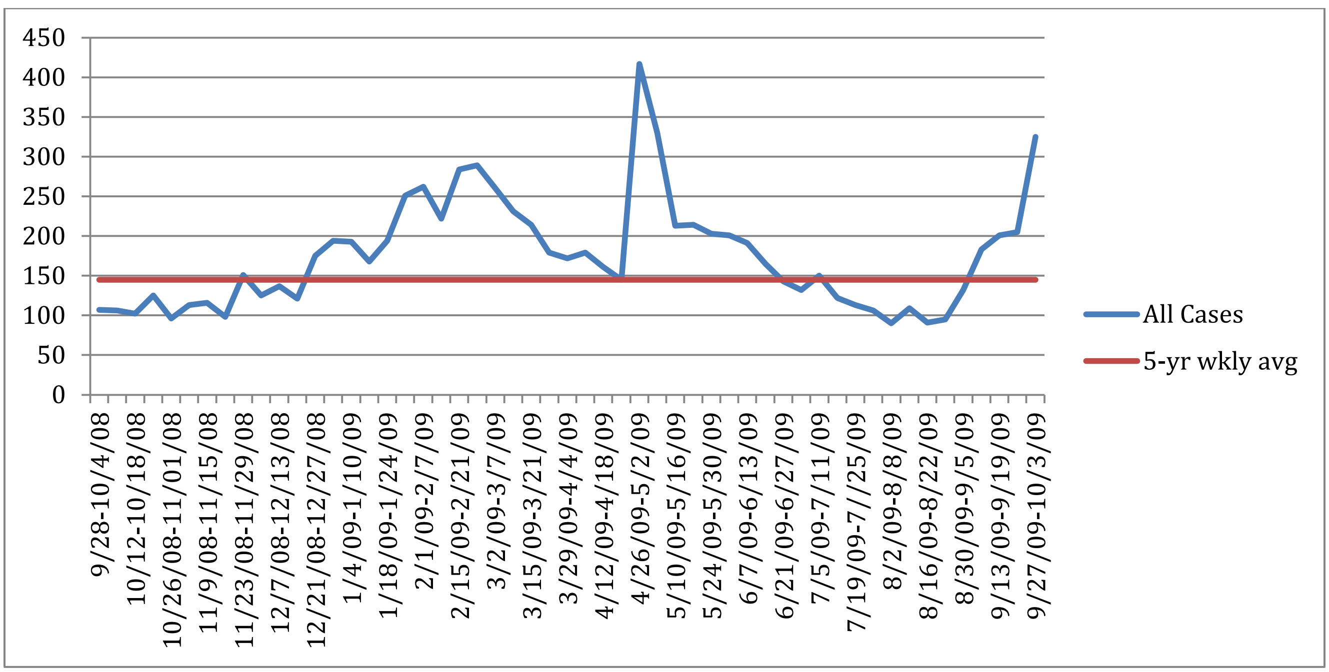


Table 1

In the table 2 below we examine the percentage of flu cases for Oct. 1, 2008 – December 31, 2009. We also averaged the five year percentage of flu hospital visits for comparison purpose. We continued through the end of 2009 since there was another spike in cases at the end of September 2009 (see table 2). We extended the one year examination period to more adequately assess the second ILI outbreak in late September 2009 to examine the duration of the outbreak. The average 5-year patterns for cases of ILI shows a similar, though higher outbreak trend. The five year average number of weekly cases also illustrates the earlier than average second outbreak. The data also appears to confirm the aberration of the April 26, 2009 spike.

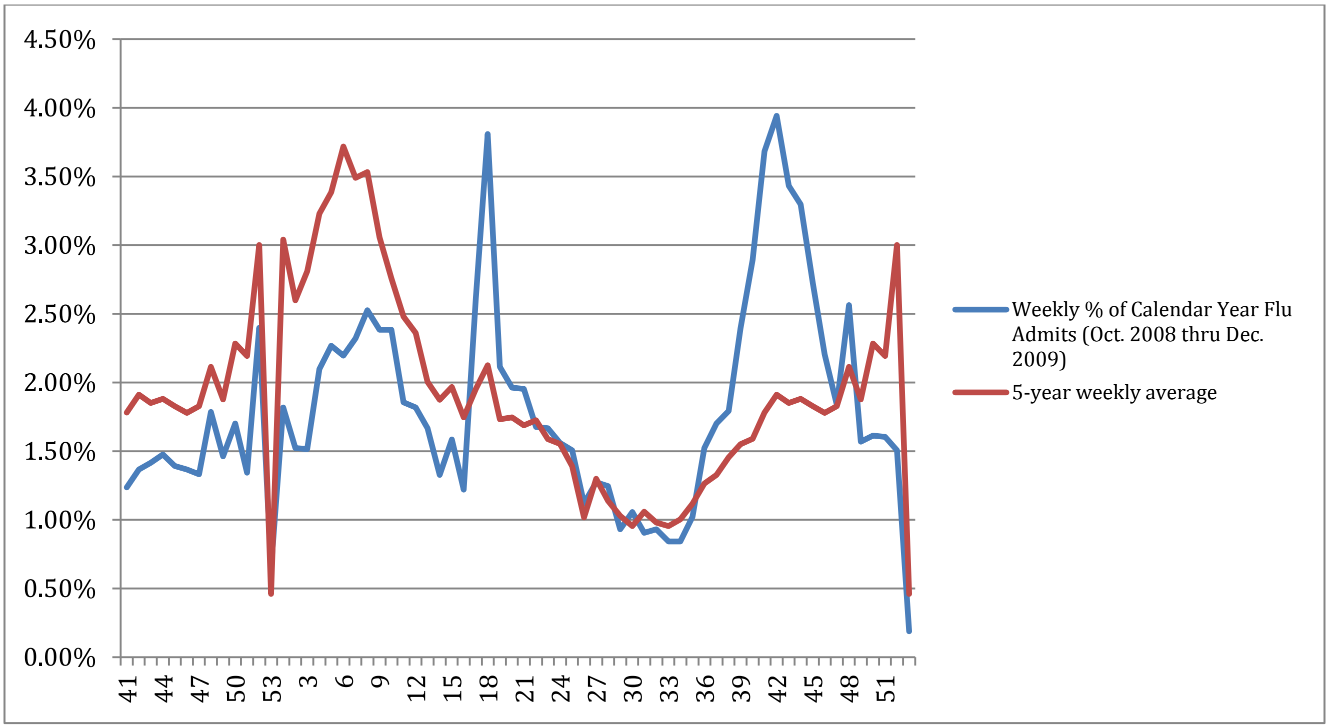


Table 2

Appendix B: Hadoop Cluster Operation

Startup/Shutdown

Startup/shutdown	1.	Press the start-up button(s) on each machine in the cluster and allow linux to complete in start up. NOTE: All machines must be running for Hadoop to work properly. *Both MySQL and the file share into the sim directory should start automatically.*
	2.	Log on Host: q000q Username: qq Password: qq *NOTE: We assume here that you have set up an entry in the local hosts file. If you're working from a Win7 machine the host file is located at C:\Windows\System32\drivers\etc\hosts See hosts file below.*
	3.	Start Hadoop qq@q000q:~$ ./start *NOTE: Give Hadoop three - five minutes to fully start.*
	4.	Stop Hadoop qq@q000q:~$ ./stop
	5.	Shutdown cluster qq@q000q:~$ ./shutdown –h *NOTE: From here you must press the start-up buttons to get the cluster going again.*
	6.	Bounce cluster qq@q000q:~$ ./shutdown –r *NOTE: Bounce = restart. Stop and starts all machines.*

Administering Hadoop

Administering Hadoop	Hadoop has three web pages that are helpful to the administrator: 1. Name node 2. Map/Reduce Administration 3. Task tracker Status
	Name Node page is accessed using a web browser. Enter: http://q000q:50070/dfshealth.jsp *NOTE: From here you can browse the hadoop file system and view the log files.*
	Map/Reduce Administration page is accessed using a web browser. Enter: http://q000q:50030/jobtracker.jsp *NOTE: This is useful for monitoring the progress of map/reduce jobs.*
	Task tracker Status page accessed using a web browser. Enter: http://q000q:50060/tasktracker.jsp *NOTE: I've never found this page to be useful.*

Running a job

Running a job	The script ./go is used to run a Hadoop job. There are two versions of it (1) can be found in q000q:/home/qq/go and (2) the other can be found in q000q:/home/sim/go.
	q000q:/home/qq/go is more generic in that it can run any M/R (Map/Reduce) job that has been assembled into a jar file. qq@q000q:~$ ./go stage.jar -s 0506LMIDBASE -r run1 -y 56 –l **stage.jar** is the complete set of biomobility M/R jobs assembled into one jar. It must reside in the /home/qq directory. **-s** specifies the name of the scenario to run. This name must match a directory in /home/sim/biomobility. The matching directory must contain a scenario.xmi file. **-r** specifies the name of the run. This name must match a directory in /home/sim/biomobility/<scenario name> The matching directory must contain a conf.xmi file. **-y** specifies the two digit flu season code. E.g. –y 56 = 2005-2006 flu season. **-l** Tells the job to copy the final files into a local directory.
	q000q:/home/sim/go is can only run the stage.jar file. qq@q000q:~$ ./go -s 0506LMIDBASE -r run1 -y 56 –l *NOTE: stage.jar is not specified in this version of the command. All other parameters remain the same as the above.*
flu season codes	**56** = 2005-2006 flu season. **67** = 2006-2007 flu season. **78** = 2007-2008 flu season. **89** = 2008-2009 flu season. **910** = 2009-2010 flu season.

Hosts file

Hosts file	Windows, Linux, Mac, and Unix all have what is known as a hosts file. A hosts file contains entries that cross reference Windows 7 keeps its file at C:\Windows\System32\drivers\etc\hosts Linux keeps its hosts file at /etc/hosts. Changing the hosts file requires sudo privleges. See sudo below. Making entries in the local (client machine's) hosts file is a more convenient way to address machines in the cluster. Example of q000q:/etc/hosts fe00::0 ip6-localnet ff00::0 ip6-mcastprefix ff02::1 ip6-allnodes ff02::2 ip6-allrouters 192.168.40.160 q000q

	192.168.40.161 q001q 192.168.40.162 q002q 192.168.40.163 q003q 192.168.40.164 q004q 192.168.40.2 bioserver More on hosts. http://en.wikipedia.org/wiki/Hosts_%28file%29
sudo	The user qq can run sudo prefixed commands the sim user cannot. Article on sudo: http://en.wikipedia.org/wiki/Sudo
biomobility directory	The directory /home/sim/biomobility is essential to the running of simulator jobs. If the –l flag is used final output files are copied out of hadoop into this directory. The basic structure is: /home/sim/biomobility/<scenario name>/<run name> The following files are require to be present. /home/sim/biomobility/<scenario name>/scenario.xmi /home/sim/biomobility/<scenario name>/<run name>/config.xmi The following files are output if the –l flag is used. /home/sim/biomobility/<scenario name>/<run name>/epistate.xmi /home/sim/biomobility/<scenario name>/<run name>/iostate.xmi qq@q000q:~$ ls /home/sim/biomobility 05-06 0607IMIN 08-09 0910IMIDBASE 0506AMID 0607LMAX 0809AMID 0910IMIN 0506IMAX 0607LMID 0809IMAX 0910LMAX 0506IMID 0607LMIDBASE 0809IMID 0910LMID 0506IMIDBASE 0607LMIN 0809IMIDBASE 0910LMIDBASE 0506IMIN 07-08 0809IMIN 0910LMIN 0506LMAX 0708AMID 0809LMAX 56crmid 0506LMID 0708IMAX 0809LMID 67 0506LMIDBASE 0708IMID 0809LMIDBASE 78 0506LMIN 0708IMIDBASE 0809LMIN 89 06-07 0708IMIN 0910 baseline 0607AMID 0708LMAX 09-10 EPI BASELINE SCENARIOS 0607IMAX 0708LMID 0910AMID resources 0607IMID 0708LMIDBASE 0910IMAX 0607IMIDBASE 0708LMIN 0910IMID
/home/sim directory	The directory /home/sim is mappable by a windows client. It contains the aforementioned biomobility directory.
scenario.xmi	<?xml version="1.0" encoding="UTF-8"?> <scenario:Scenario xmi:version="2.0" xmlns:xmi="http://www.omg.org/XMI" xmlns:scenario="qq.mr.scenario.xsd" begin="2005-10-09T00:00:00" end="2006-04-01T00:00:00" stepSize="1440" stepBack="20160" stayHome="0.25" diseaseOfInterest="Influenza-E1-I-5-F.4" airportOfInterest="LAS"

	averageStayDuration="5040" dataSource="all-2005-2006.data" title="56crmid"> <outbreak> <locale title="Boston" contactRate="0.56" population="609023"/> <y0Primes/> </outbreak> <outbreak> <locale title="Philadelphia" contactRate="0.56" population="1400000"/> <y0Primes/> </outbreak> <localeOfInterest title="Las Vegas" contactRate="1.0" population="2000000"/> <nationalY0Prime> <primeSet key="2005-10-08"> <values> <value value="0.987627265394084" name="S"/> <value value="0.00582" name="E"/> <value value="0.0060" name="I"/> <value value="5.52734605915761E-4" name="R"/> </values> </primeSet> Many more prime sets… </nationalY0Prime> </scenario:Scenario>

Appendix C: Required Simulator File Descriptions

This file must reside in /home/sim/biomobility/<scenario name>/scenario.xmi		
Element name	Attribute name	Explanation
scenario:Scenario	begin	Date upon which the simulation is to begin. This not necessarily the first date in the input file. The simulator will skip all input records that are prior to this date.
	end	Date upon which the simulation is to end. This not necessarily the last date in the input file. The simulator will skip all input records that are after to this date.
	stepSize	Expresses the degree of granularity the simulator uses with regard to time as expressed in minutes.
	stepBack	A span of time reaching back to before the begin date. The simulator uses this time span to ramp up the visitor population to a desired level for processing. The value is expressed in minutes.
	stayHome	A percentage (e.g. .25) by which the simulator will reduce the infectious population. It is assumed that sick people will elect not to travel.
	diseaseOfInterest	The disease profile to use in processing this scenario. It must match an entry in the resources/disease.xmi file or an error is thrown.
	airportOfInterest	The airport code of the locale of interest. Not used, deprecated
	averageStayDuration	Not used, deprecated
	dataSource	The essential input file. This file is read from the directory /hdfs/sourcedata/<dataSource file name>
	title	Name of the scenario. Must match the name of the directory in which the scenario file resides.
outbreak		Defines a local where an outbreak takes place.
locale	title	Name of the outbreak locale.
	contactRate	ContactRate for the outbreak locale.
	population	Population of the outbreak locale. Not used, deprecated
y0Primes		Defines the set y0prime values for the outbreak locale. This element's contents are structured that same as is nationalY0Prime below.
localeOfInterest	title	Locale that is the center of the simulation. E.g. Las Vegas.
	contactRate	ContactRate for the locale.
	population	Population of the locale. Not used, deprecated
nationalY0Prime		Defines the set y0prime values for the nation. Data values are taken from the CDC.
primeSet	key	The date for which the values are applicable.
values		
value	name	Name of the value. i.e. S,E,I, or R.
	value	The percentage of the population that is part of this stage at this time.

Scenario.xmi Description

Disease.xmi Description

This file must reside in /home/sim/biomobility/resources/disease.xmi		
Element name	Attribute name	Explanation
disease:Diseases		
disease	title	Name of the disease. Must match the diseaseoOfInterest in the scenario.xmi file.
	force	Force of infection
stages	code	S, E, I, or R
	title	Susceptible, Exposed, Infected, or Resistant. Must correlate with the code.
	ordinal	Order of progression through the disease starting with zero.
	duration	The length of time expressed in minutes that one remins at this stage. The value -1 indicates an indefinite period of time.
	susceptible	True or false is the person susceptible that this stage.
	infected	True or false is the person infected that this stage.
	infectious	True or false is the person infectious that this stage.

Conf.xmi Description

This file must reside in /home/sim/biomobility/<scenario name>/<run name>/conf.xmi		
Element name	Attribute name	Explanation
conf:Configuration		
Slim, round, prime, split, progress, contact, depart, consolidate, and ioconsolidate *NOTE: for best results, do not modify this file. Make a copy if you need a new one.*	gonogo	True/false indicates whether or not to run this stage.
	i	Always false
	o	Always false
	w	Always true

In the directory /home/sim/biomobility, any scenario whose name follows the pattern 0506AMID or 0506IMIDBASE is a production scenario. The others are not.
In the directory /home/sim,

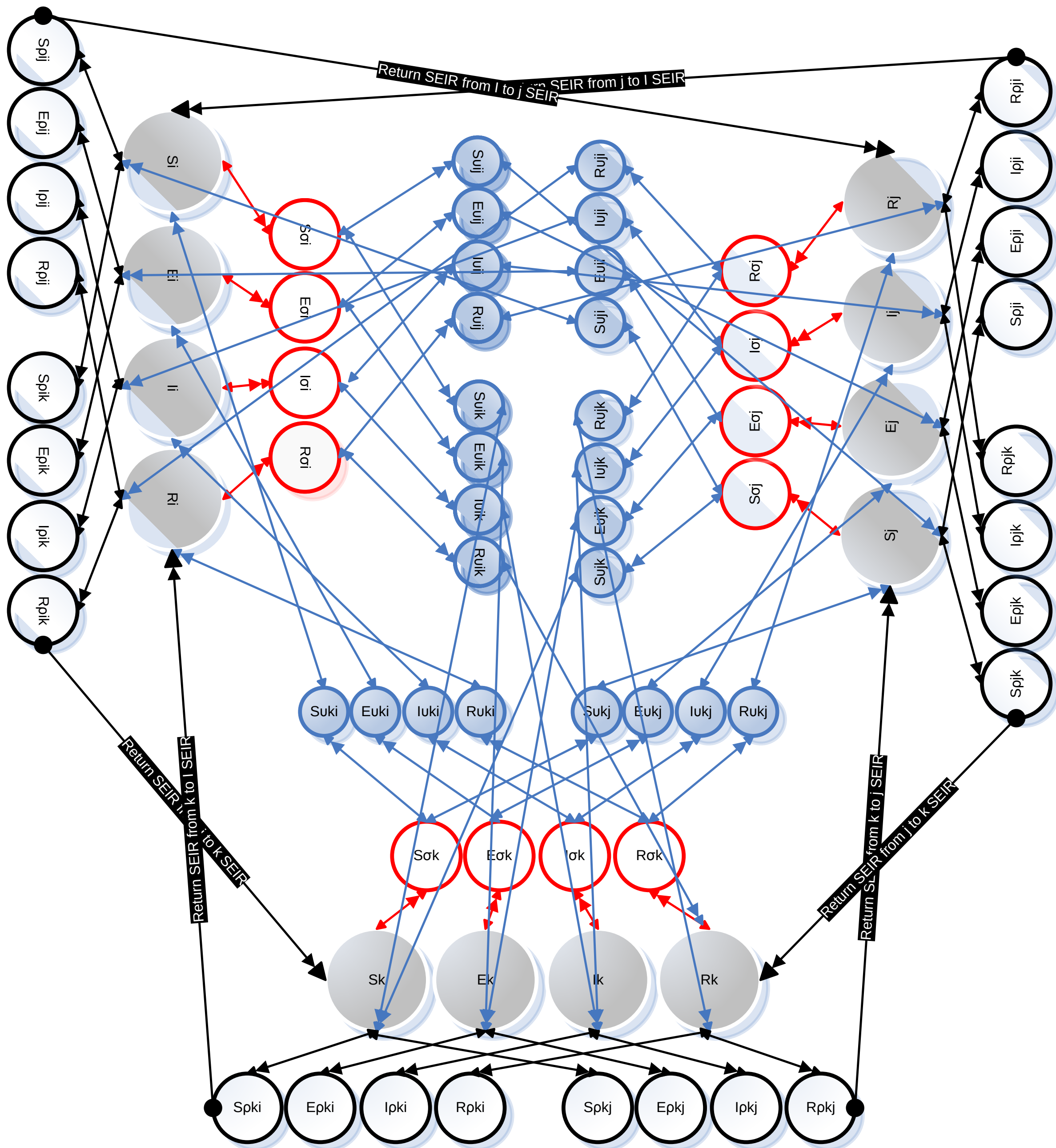
iostate-mindx.xlsx	A spreadsheet of the min, mid and max infectious departures. Flights only.
Flights.png	An image of the US showing the flights.

In the directory /home/qq, the following are scripts that are potentially useful.

go	Runs a hadoop job.
H	Helps manipulate the hdfs (Hadoop Distributed File System). e.g. –copyFrom Local <path to local disk> <path to hdfs> copies a file into the hdfs. Full list of commands: http://hadoop.apache.org/common/docs/r0.17.1/hdfs_shell.html
clear	Deletes all the hadoop logs on all nodes in the cluster.
kill	Kills a hadoop job. Requires the job number of the job you're trying to kill as a parameter. Job numbers are output when a job starts.
start	Starts hadoop; all nodes.
stop	Stops hadoop; all nodes.
shutdown	Shutsdown the cluster. Requires either and –h or –r flag. –h is for halt. –r is for restart.
envars	Contains all environment variables. Called by some of the other commands. No called directly.
slaveloop	Iterates through the nodes in the cluster. Called by some of the other commands. No called directly.

Appendix D: Miscellaneous

State Space Model for SEIR in Three Cities



Legend

Grey = City
Red = Leave
Blue = Visit
Black = Return

Variables

Si, Sj, Sk	Susceptible population in each city	Initial Value = 85% of population
Ei, Ej, Ek	Exposed and latent population in each city	Initial Value = 10% of population
Ii, Ij, Ik	Infectious population in each city	Initial Value = 5% of population
Ri, Rj, Rk	Removed population in each city	Initial Value = 0% of population
Ni, Nj, Nk	Total population in each city	Initial Value = 100% of population

Constants

Sigma	σ	Leave Rate = 20%
Upsilon	υ	Visit Rate = 10%
Rho	ρ	Return Rate = 10%
Kappa	κ	Contact Rate = 20/day
Beta	β	Transmission Rate = 2 new infections for every infective per day
Epsilon	ε	Infection Rate = 10%
Gamma	γ	Recovery Rate = 5 days

Equations

Ni = Si + Ei + Ii +Ri
Nj = Sj + Ej + Ij +Rj
Nk =Sk + Ek + Ik +Rk

System of Differential Equations

/* city I   SEIR */
dSi/dt = Sρki + Sρkj – Sσi – β * Si *Ii/Ni
dEi/dt = Eρki + Eρkj – Eσi + κ* β * Si *Ii/Ni – εEi
dIi/dt = Iρki + Iρkj – Iσi + εEi – γIi
dRi/dt = Rρki + Rρkj - Rσi + γIi

/* city I leave equations   */
dSσi/dt = σ * Si
dEσi/dt = σ * Ei
dIσi/dt = σ * Ii
dRσi/dt = σ * Ri

/* city I leaving visitor, visit distribution equations   */
/* to city j */
dSυij/dt = υ * Sσi
dEυij/dt = υ * Eσi
dIυij/dt = υ * Iσi
dRυij/dt = υ * Rσi
/* to city k */
dSυik/dt = υ * Sσi
dEυik/dt = υ * Eσi
dIυik/dt =   υ * Iσi
dRυik/dt = υ * Rσi

```
/* city I returning visitor equations */
/* from city j */
dSρji/dt =  ρ * Sj
dEρji/dt =  ρ * Ej
dIρji/dt =  ρ * Ij
dRρji/dt =  ρ * Rj

/* from city k */
dSρki/dt =  ρ * Sk
dEρki/dt =  ρ * Ek
dIρki/dt =  ρ * Ik
dRρki/dt =  ρ * Rk

/* city J   SEIR */
dSj/dt = Sρkj + Sρij – Sσj – β * Sj *Ij/Nj
dEj/dt = Eρkj + Eρij – Eσj + κ* β * Sj *Ij/Nj – εEj
dIj/dt = Iρkj + Iρjj – Iσj + εEj – γIj
dRj/dt = Rρkj + Rρij - Rσi + γIi
/*  city I leave equations   */
dSσj/dt = σ * Sj
dEσj/dt =  σ * Ej
dIσj/dt =  σ * Ij
dRσj/dt =  σ * Rj

/*  city J leaving visitor, visit distribution equations   */
/* to city i */
dSυji/dt =  υ * Sσj
dEυji/dt =  υ * Eσj
dIυji/dt =  υ * Iσj
dRυji/dt =  υ * Rσj

/* to city k */
dSυjk/dt =  υ * Sσj
dEυjk/dt =  υ * Eσj
dIυjk/dt =   υ * Iσj
dRυjk/dt =  υ * Rσj

/* city J returning visitor equations */
/* from city i */
dSρij/dt =  ρ * Si
dEρij/dt =  ρ * Ei
dIρij/dt =  ρ * Ii
dRρij/dt =  ρ * Ri

/* from city k */
dSρkj/dt =  ρ * Sk
```

$dE\rho kj/dt = \rho * Ek$
$dI\rho kj/dt = \rho * Ik$
$dR\rho kj/dt = \rho * Rk$

/* city k SEIR */
$dSk/dt = S\rho ik + S\rho jk – S\sigma k – \beta * Sk *Ik/Nk$
$dEk/dt = E\rho ik + E\rho jk – E\sigma k + \kappa * \beta * Sk *Ik/Nk – \varepsilon Ek$
$dIk/dt = I\rho ik + I\rho jk – I\sigma k + \varepsilon Ek – \gamma Ik$
$dRk/dt = R\rho ik + R\rho jk - R\sigma k + \gamma Ik$

/* city k leave equations */
$dS\sigma k/dt = \sigma * Sk$
$dE\sigma k/dt = \sigma * Ek$
$dI\sigma k/dt = \sigma * Ik$
$dR\sigma k/dt = \sigma * Rk$

/* city k leaving visitor, visit distribution equations */
/* to city i */
$dS\upsilon ki/dt = \upsilon * S\sigma k$
$dE\upsilon ki/dt = \upsilon * E\sigma k$
$dI\upsilon ki/dt = \upsilon * I\sigma k$
$dR\upsilon ki/dt = \upsilon * R\sigma k$

/* to city j */
$dS\upsilon kj/dt = \upsilon * S\sigma k$
$dE\upsilon kj/dt = \upsilon * E\sigma k$
$dI\upsilon kj/dt = \upsilon * I\sigma k$
$dR\upsilon kj/dt = \upsilon * R\sigma k$

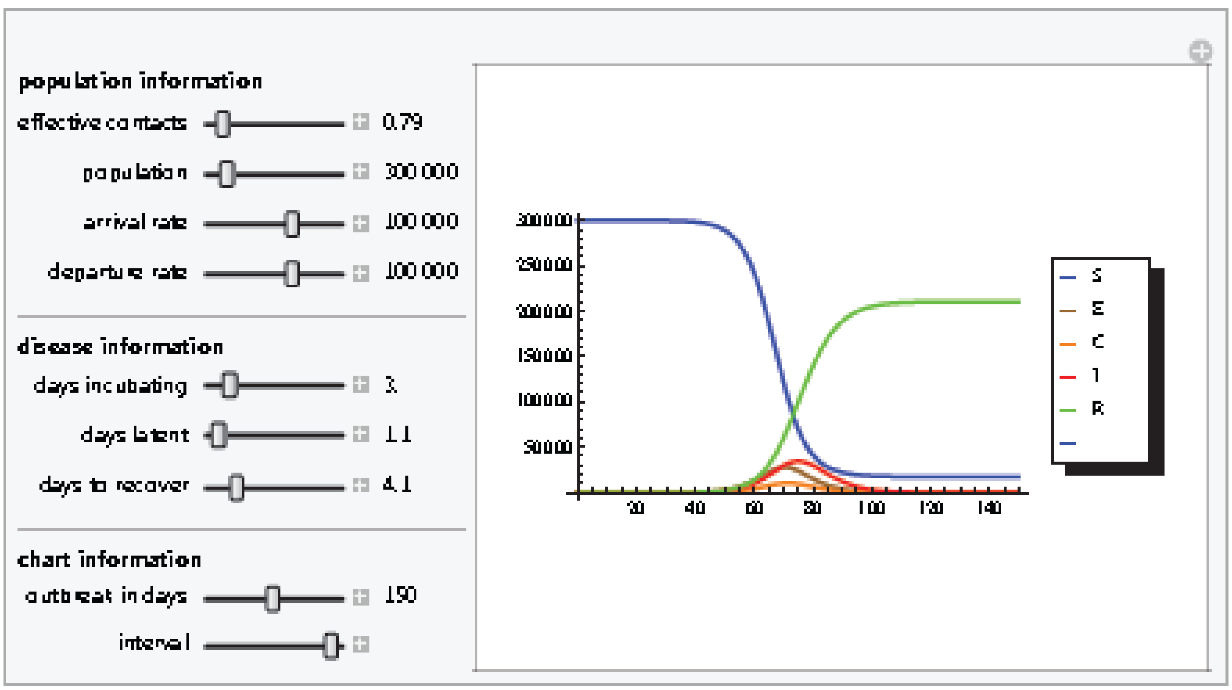
/* city k returning visitor equations */
/* from city i */
$dS\rho ik/dt = \rho * Si$
$dE\rho ik/dt = \rho * Ei$
$dI\rho ik/dt = \rho * Ii$
$dR\rho ik/dt = \rho * Ri$

/* from city j */
$dS\rho jk/dt = \rho * Sj$
$dE\rho jk/dt = \rho * Ej$
$dI\rho jk/dt = \rho * Ij$
$dR\rho ki/dt = \rho * Rj$

Compartment Model with Carrier Status to Reflect Subclinical Viral Shedding

```
SEIR Eqns V10 test.nb
In[1]:=

 This is an SECIR model with a compartment for CARRIER status between
  E and I to reflect infectivity before the onset of symptoms

   uses proportional mixing and inverse of days latent and duration
   arrivals and departures evenly split

    beta = effective contacts or transmission coefficient or replacement rate
    sigma = incubating
      lamda = duration of carrier infectivity
      gamma = duration of symptoms

       SEIR equations:
           ds dt =    bsi  n    .25 a     .25 d
           de dt =   bsi  n    1 Σ e    .25 a     .25 d
           dc dt =  1 Σ e    1 Λ  i    .25 a     .25 d
            di dt =   1 Σ c    1 Γ  i    .25 a     .25 d
            dr dt =   1 Γ i    .25 a     .25 d

 Needs["PlotLegends`"]

 Manipulate[
  Plot[
   Evaluate[
    {s[t], e[t], c[t], i[t], r[t]} . NDSolve[{s'[t]     b  s[t]  i[t]   popln
          b   s[t]  c[t]   popln     .25   a     .25   d , s[1]   popln   1 ,
      e'[t]     b   s[t]  i[t]   popln     1    Σ    e[t]    .25   a     .25   d  ,
      e[1]   1.0,
      c'[t]      1   Σ   e[t]    1   Λ   c[t]     .25   a     .25   d  , c[1]   0,
      i'[t]     1   Λ   c[t]    1   Γ   i[t]     .25   a     .25   d  , i[1]   0,
      r'[t]    1   Γ   i[t]     .25   a     .25   d  , r[1]   0},
     {s , e, c, i, r}, {t, 0, 150}]    end NDSolve

   ]  end Evaluate  ,
     EvaluationMonitor:-Print["S = ",s[t]" E = ",e[t] "I = ",i[t]" R =  ",r[t]]
   {t, 0.1, tmax},
   PlotStyle �
    {{Blue, Thick},  {Brown, Thick} , {Orange,  Thick},  {Red, Thick}, { Green,  Thick}} ,

   PlotLegend � {"S",  "E", "C", "I", "R"}, LegendPosition � {1.1,  0.4}]


 (* end Plot *)

 (* manipulation controls *)
 , Delimiter
 , Style["population information", Bold]
 , {{b, 0.79, "effective contacts"},
  0, 20, 0.01, ImageSize � Tiny, Appearance � "Labeled"}
  , {{popln, 300 000, "population"}, 150 000, 2 000 000, 1000,
  ImageSize � Tiny, Appearance � "Labeled"}
```

```
 , {{a, 100 000, "arrival rate"}, 50, 150 000, 1000,
 ImageSize � Tiny, Appearance � "Labeled"}
 , {{d, 100 000, "departure rate"}, 50, 150 000, 1000,
 ImageSize � Tiny, Appearance � "Labeled"}

, Delimiter
, Style["disease information", Bold]
, {{Σ, 3.0, "days incubating"}, 1, 20, 0.05,  ImageSize � Tiny, Appearance � "Labeled"}
, {{Λ, 1.1, "days latent"}, 1, 20, 0.05,  ImageSize � Tiny, Appearance � "Labeled"}
, {{Γ, 4.1, "days to recover"}, 1, 20, 0.05, ImageSize � Tiny, Appearance � "Labeled"}

, Delimiter
, Style["chart information", Bold]
, {{tmax, 150, "outbreak in days"},
 0.2, 300, 0.1,  ImageSize � Tiny, Appearance � "Labeled"}
, {{vint, 1, "interval"}, 0.05, 1, 0.01, ImageSize � Tiny}
, ControlPlacement � Left,
TrackedSymbols � Manipulate, AutorunSequencing � {1, 2, 3, 4, 5}
] (* end Manipulate *)
```



```
H*
Rules
DL >= 1.5 < 8
DL:DR = from 1:0.7 to 1:1.2
*L
```

SEIR Eqns v10 test SECIR.nb **3**